

AD _____

Award Number: DAMD17-03-1-0023

TITLE: Intensity Modulated Radiation Treatment of Prostate
Cancer Guided by High Field MR Spectroscopic Imaging

PRINCIPAL INVESTIGATOR: Lei Xing, Ph.D.

CONTRACTING ORGANIZATION: Leland Stanford Junior University
Stanford, CA 94305-4125

REPORT DATE: May 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20060110 124

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY
(Leave blank)**2. REPORT DATE**
May 2005**3. REPORT TYPE AND DATES COVERED**
Annual (1 May 04 - 30 Apr 05)**4. TITLE AND SUBTITLE**

Intensity Modulated Radiation Treatment of Prostate Cancer Guided by High Field MR Spectroscopic Imaging

5. FUNDING NUMBERS

DAMD17-03-1-0023

6. AUTHOR(S)

Lei Xing, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)Leland Stanford Junior University
Stanford, CA 94305-4125

E-Mail: lei@reyes.stanford.edu

8. PERFORMING ORGANIZATION REPORT NUMBER**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012**10. SPONSORING / MONITORING AGENCY REPORT NUMBER****11. SUPPLEMENTARY NOTES****12a. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE**13. ABSTRACT (Maximum 200 Words)**

With the development of computer controlled multileaf collimators, IMRT now provides unprecedented means for us to deliver 3D-dose distributions with sub-centimeter resolution. Simultaneously, MRSI is capable of noninvasively providing critically needed 3D metabolic and physiological information. Together with anatomical CT/MRI, functional imaging affords valuable 4D data (3D structural plus 1D metabolic). A timely issue is how to integrate the metabolic/functional data into IMRT planning to improve clinical prostate cancer management. Toward establishing biologically conformal radiation therapy for prostate treatment, we have carried out a series of studies and developments and have successfully accomplished a number of important milestones of the proposed project. These include (i) refined the endorectal coil-based 3T prostate MRSI technique for prostate imaging; (ii) developed a narrow band technique for deformable image registration of endorectal coil-based MRI/MRSI and treatment planning CT images; (iii) developed a novel inverse planning system that is capable of incorporating prior clinical knowledge; (iv) established an adaptive technique for producing optimal voxel-dependent penalty scheme for prostate IMRT planning; (v) proposed a biological model-based BCRT inverse planning technique and laid technique foundation for BCRT. It is expected these tools will greatly facilitate the imaging, planning, delivery, and quality assurance of MRSI-guided prostate treatment.

14. SUBJECT TERMS

Prostate Cancer

15. NUMBER OF PAGES

186

16. PRICE CODE**17. SECURITY CLASSIFICATION OF REPORT**

Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT

Unclassified

20. LIMITATION OF ABSTRACT

Unlimited

Table of Contents

1.	Cover page	1
2.	SF 298	2
3.	Table of Contents	3
3.	Introduction	4
3.	Body	4
4.	Key Research Accomplishments	8
5.	Reportable Outcomes	9
6.	Conclusions	10
8.	References	10
14.	Appendices	12

I. INTRODUCTION

This Idea Award (DAMD17-03-1-0023, entitled "Intensity Modulated Radiation Treatment of Prostate Cancer Guided by High Field MR Spectroscopic Imaging") was awarded to the principal investigator (PI) for the period of May 1, 2003—April 30, 2006. This is the annual report for the second funding period (May 1, 2004 –April 30, 2005). The goal of this project is to establish biologically conformal - as opposed to anatomically conformal - IMRT as a viable modality through integration with 3T MRSI imaging to more effectively kill prostate tumor cells. The underlying hypothesis driving this work is that the MRSI-guided IMRT will provide substantially improved dose distributions required to achieve greater local tumor control while maintaining, or reducing, complications to sensitive structures. The specific aims of the project are: (1) To establish a robust procedure for registering and mapping of MR spectroscopic data to CT/MRI images for prostate irradiation. (2) To develop an inverse planning system for MRSI-guided IMRT prostate treatment and demonstrate the feasibility of concurrent dose escalation to intraprostatic lesion(s) through a set of phantom studies and at least two previously treated prostate cases who had undergone CT/MRSI scans. Under the generous support from the U.S. Army Medical Research and Materiel Command (AMRMC), the PI has contributed significantly to prostate cancer research by applying physics and engineering knowledge to prostate cancer research. A number of significant conference abstracts and refereed papers have been resulted from the support. The preliminary data obtained under the support of the grant has also enabled the PI to start new research initiatives and significantly advanced his academic career. In this report, the past year's research activities of the PI are highlighted.

II. RESEARCH AND ACCOMPLISHMENTS

Prostate Intensity modulated radiation therapy (IMRT) is an image-guided process whose success depends critically on the imaging modality used for treatment planning and the level of integration of the available information. In current clinical practice, IMRT prostate treatment planning, performed under the guidance of CT or MRI, is aimed at achieving uniform dose distribution to the entire prostate volume. While tumor biology plays a crucial role in the treatment outcome, neither MRI nor CT provides biological information¹⁻⁵. Recent advancement of MR spectroscopic imaging (MRSI)⁶⁻¹⁰, especially high field (3 Tesla) MRSI, now makes possible to effectively distinguish between regions of cancer and normal prostatic epithelium with clinically acceptable resolution ($\sim 3\text{mm}$)^{10, 11}. Coupled with the technical capability of IMRT (intensity modulated radiation therapy) and IGRT (image-guided radiation therapy) for delivering customized 3D dose distributions, an important and timely question to ask is whether the metabolic

information derived from MRSI can be used for more objective tumor delineation and for guiding IMRT treatment planning.

Toward establishing biologically conformal IMRT for prostate cancer treatment, we have done substantial work in the past year in integration of MRSI imaging into IMRT treatment planning and in improving the existing prostate IMRT planning techniques. The research is sorted into five different subjects and summarized below.

Development of 3T MRSI techniques: Our technical development efforts for enhancing 3T MRSI over the past year have focused on developing a prostate MRSI sequence for imaging choline and citrate has been developed and tested on several subjects. The method is based on a 2D J-resolved acquisition in order to better identify the citrate and choline resonances. These Rf pulses, consisting of a match 90-180 spectral spatial pair, are optimized for choline imaging while minimizing peak Rf power constraints that typically limit 3T body MRI¹⁰. Our ultimate goal is the development and systematic evaluation of quantitative 3.0 T volumetric spiral 1H-MRSI pulse sequences optimized for high spatial resolution, minimum acquisition time, and maximum information content to meet the future needs of high-field spectroscopic imaging and provide improved treatment target definition and more accurate treatment monitoring.

Image registration using narrow band method: Image registration is the determination of a geometrical transformation that aligns points in one view of an object with corresponding points in another view of that object or another object. In the last funding period we have developed a the *thin-plate splines* (TPS) method^{12,13} to deal with the registration of endorectal MRI/MRSI and verified the accuracy of the registration by phantom and patient studies. The technique provide a clinically useful solution to register endorectal probe based MRI/MRSI and CT. However, a set of control point pairs must be selected manually on the model and reference images, which is a tedious process and often lead to registration inaccuracy. In this funding period, we have made effort to maximally automate the registration process. A narrow band registration, which is a hybrid method combining the advantages of pixel-based and distance-based registration techniques, has been developed to directly register ER-based MRI/MRSI with CT¹⁴. The normalized correlation between the two input images for registration was used as the metric and the calculation was restricted to those points contained in the narrow bands around the user-delineated structures. The narrow band method is inherently efficient because of the use of *a priori* information of the meaningful contour data. The ability of the algorithm to recover the original image was assessed using checkerboard graph. The mapping of ER-based MRI onto treatment planning CT images was carried out

for two clinical cases and the performance of the registration was evaluated. It was shown that the technique can restore an MR image from intentionally introduced deformations with an accuracy of ~2 mm. Application of the technique to two clinical prostate MRI/CT registrations indicated that it is capable of producing clinically sensible mapping. The whole registration procedure for a complete 3D study (containing 256x256x64 voxels) took less than 15 minutes on a standard PC with much less user interaction as compared with the TPS method. This study has recently published in the International Journal of Radiation Oncology, Biology, and Physics (one of the figures from the study was selected to be the cover picture of the June issue of the journal) ¹⁴.

Image registration with auto-mapped control points: An innovative registration method capable of auto-select the control point pairs has been developed to fully automate the general image registration process¹⁵. Many image registration algorithms rely on the use of homologous control points on the two input image sets to be registered. In reality, the interactive identification of the control points on both images is tedious, difficult and often a source of error. In this work we proposed a technique that allows us to automate the selection of control points for both rigid and deformable image registrations. The registration of two images in our approach proceeds in two steps. First, a number of small control regions having distinct anatomical features are identified on the model image. Instead of attempting to find the correspondences of the control regions in the reference image through user interaction, in the proposed method, each of the control regions is mapped to the corresponding parts of the reference image by using an automated image registration algorithm. The conventional automated image registration algorithm is then used to complete the image registration process with the guidance of auto-determined control points. The performance of the two-step or registration-in-registration method was examined by MRI-CT registration in a variety of situations. Our results indicated that the approach is robust and has great potential for clinical MRI/MRSI-CT registration.

Incorporation of clinical outcome data to guide inverse planning. IMRT plan ranking is, to a certain degree, subjective and the solution depends strongly the selection of models for inverse planning. We have developed a formalism to integrate existing clinical endpoint data to better guide the inverse planning optimization process¹⁶. The new formalism is biologically more sensible yet clinically practical for functional image-guided IMRT planning. In this formalism we propose to characterize the dose-volume status of a structure by using the concept of effective volume in the voxel domain and constructed an objective function with incorporation of the volumetric information. Compared with the conventional

inverse planning technique, we found that, for the same target dose coverage, the critical structure sparing was substantially improved for all testing cases.

IMRT inverse planning with adaptively determined voxel-dependent importance factors: Voxels within a target or a sensitive structure volume are generally not equivalent in achieving their dosimetric goals in IMRT planning. Inverse planning objective function should not only balance the competing objectives of different structures but also that of the individual voxels in various structures. This issue becomes especially important when dealing with MRSI-guided IMRT planning because a spatially non-uniform dose distribution is often desired. We have proposed an approach to improve IMRT treatment plan by purposely modulating the penalty on individual voxel level. While it is permissible for each voxel to have its own importance value, a challenging problem is how to obtain a sensible set of importance factors with a manageable amount of computing. This issue was resolved by using an adaptive algorithm, in which, instead of letting the values of voxel dependent importance to vary freely during the search process, they were tied to the local radiation doses through a heuristically constructed relation. The new planning tool was applied to study a hypothetical phantom case and a prostate case and comparison of the results with that obtained using conventional inverse planning technique with structure specific importance factors indicated that the dose distributions from the conventional inverse planning are at best sub-optimal and can be significantly improved with the help of the proposed non-uniform penalty scheme. We have shown that an inverse planning framework can lead to significantly improved IMRT treatment plans that would otherwise be unattainable. The work has been Published in *Medical Physics Journal*¹⁷.

Inverse planning for biologically conformal IMRT: Recent progress in biological imaging such as MRSI and PET is making the mapping of this distribution increasingly possible. But an inverse planning system that allows us to utilize the spatial biology distribution does not exist. In this work we have established a theoretical framework to quantitatively incorporate the spatial biology data into IMRT inverse planning⁵. In order to implement this, we first derive a general formula for determining the desired dose to each tumor voxel for a known biology distribution of the tumor based on a linear-quadratic (LQ) model. The desired target dose distribution is then used as the prescription for inverse planning. An objective function with the voxel-dependent prescription is constructed with incorporation of the nonuniform dose prescription. Our calculations revealed that it is technically feasible to produce deliberately nonuniform dose distributions with consideration of biological information. Compared with the conventional dose escalation schemes, the new technique is capable of generating biologically

conformal IMRT plans that significantly improve the tumor control probability (TCP) while reducing or keeping the normal tissue complication probability (NTCP) at their current levels. Biologically conformal radiation therapy (BCRT) incorporates patient specific biological information and provides an outstanding opportunity for us to truly individualize radiation treatment. The proposed formalism lays a technical foundation for BCRT and allows us to maximally exploit the technical capacity of IMRT to more intelligently escalate the radiation dose.

The Idea Award for Prostate Cancer Research from US Army Medical Research and Materiel Command also provides a unique educational opportunity for training junior researcher through the participation of research activities. In this aspect, the postdoctoral fellow, Eduard Schreibmann, has been benefited greatly from the support. After obtaining his Ph.D. degree in Medical Physics from University of Patras, Greece, Dr. Schreibmann jointed my group last year. Since he came here, he has had opportunity to learn the clinical prostate treatment planning, simulation, and quality assurance and many aspects of radiation treatment of prostate cancer. He has developed an elegant and clinically practical method of beam orientation optimization for prostate IMRT. He has also done substantial work on MRSI-CT image registration. He is close to complete his postdoc training at Stanford. Given his training obtained under the support of this grant and his performance in the past two years, I expect that he will become a leading researcher in the fields of medical physics and prostate radiotherapy in the years to come.

III. KEY RESEARCH ACCOMPLISHMENTS

- Refined the endorectal coil-based 3T prostate MRSI technique for prostate imaging.
- Developed a narrow band technique for deformable image registration of endorectal coil-based MRI/MRSI and treatment planning CT images.
- Developed a novel inverse planning system that is capable of incorporating prior clinical knowledge.
- Established an adaptive technique for producing optimal voxel-dependent penalty scheme for prostate IMRT planning.
- Proposed a biological model-based BCRT inverse planning technique and laid technique foundation for BCRT.

IV. REPORTABLE OUTCOMES

The following is a list of publications resulted from the grant support. Copies of the publication materials are enclosed with this report.

Refereed publications:

1. Yong Y, Xing L, "Inverse planning with adaptively evolving voxel-dependent penalty scheme", *Medical Physics* 31, 2819-2825, 2004.
2. Lian J, Hunjan S, Daniel B, Lo A, Levin J, Cardenas C, Dumoulin C, Watkins R, Rohling K, Giaquinto R, Boyer A, Xing, L. Mapping of the Prostate in Endorectal Coil-Based MRI/MRSI and CT: A Deformable Registration and Validation Study. *Medical Physics* 31: 2711-2720, 2004.
3. Schreibmann E and Xing L: "Narrow band deformable registration of prostate MRI/MRSI and CT studies". *International Journal of Radiation Oncology, Biology, Physics* 62, 595-605, 2005.
4. Schreibmann E and Xing L, "Beam orientation class-solutions for IMRT prostate cancer treatment". *Medical Physics* 31, 2863-2870, 2004.
5. Yang Y and Xing L: "Clinical knowledge-based inverse treatment planning". *Physics in Biology and Medicine* 49, 5101-5117, 2004.
6. Kim D, Mayer D., Xing L, Daniel B, Margolis, D., Spielman D., "In vivo detection of citrate for prostate cancer at 3 Tesla", *Magnetic Resonance Imaging in Medicine*, 53, 1177-1182, 2005.
7. Xia P, Yu N, Xing L, Syn X, Verhey, L: "Investigation o a new objective function for inverse planning optimization ". *Medical Physics* 32, 920-927, 2005.
8. Yang Y and Xing L: "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution ". *Medical Physics* 32, 1473-84, 2005.
9. Schreibmann E and Xing L: "Dose-volume based ranking of incident beam directions and its utility in facilitating IMRT beam placement". *International Journal of Radiation Oncology, Biology, Physics*, submitted.
10. Schreibmann E and Xing L: "Image registration with auto-mapped control volumes". *International Journal of Radiation Oncology, Biology, Physics*. submitted

Book Chapters:

1. Xing L., Yang Y., Spielman D., Molecular/Functional Image-Guided Radiation Therapy, T. Bortfeld, R. Schmidt-Ullrich, W. de Neve (editors), Springer-Verlag Heidelberg, Berlin, in press.
2. Xing L, Wu Q, Yong Y and Boyer AL: Physics of IMRT and Inverse Treatment Planning, in Intensity Modulated Radiation Therapy: *A Clinical Perspective*, Mundt AF and Roeske JC. (editors), page 20-51, BC Decker Inc. Publisher, Hamilton, Canada, 2005.

Conference abstract:

1. Xing, L. and Spielman D, Integration of MRI/MRSI into Radiation Therapy Treatment Planning, 2005 AAPM Annual Meeting, Seattle, WA (invited talk).
2. Schreibmann E. and Xing L., Registration of prostate MRI/MRSI and CT studies using narrow band approach, 2005 AAPM Annual Meeting, Seattle, WA, submitted.
3. Schreibmann E. and Xing L., EUD-based beam orientation optimization, 2005 AAPM Annual Meeting, Seattle, WA, submitted.
4. Xing L, Levy, D. and Yang Y., Incorporating clinical outcome data into inverse treatment planning, 2005 AAPM Annual Meeting, Seattle, WA, submitted.
5. Yang Y. and Xing L, Prescription for biologically conformal radiation therapy, 2005 AAPM Annual Meeting, Seattle, WA, submitted.
6. Schreibmann E, and Xing L: "Image registration with auto-mapped control volumes". Submitted to 2005ASTRO annual meeting, Denver, CO.

7. Yang Y and Xing L: "Optimization of radiation dose-time-fractionation scheme with consideration of tumor specific biology". Submitted to 2005ASTRO annual meeting, Denver, CO.

V. CONCLUSIONS

MR Spectroscopic Imaging-guided IMRT technique is being developed for the treatment of prostate cancer. A few important milestones have been achieved toward the general goal of the project. These include (i) implemented 3T prostate MRSI technique with J-resolved pulse sequence; (ii) development of a robust image registration procedure to map endorectal coil-based MRSI data onto treatment planning CT/MRI images; (iii) experimental validation of the image registration tools; (iv) phantom measurement of MRSI; (v) improvement of currently available inverse planning algorithm for MRSI-guided prostate IMRT. Integration and further refinement of the above tools are underway. It is expected these tools will greatly facilitate the planning, delivery, and quality assurance of the MRSI-guided prostate treatment.

References:

1. Ling, C. C.; Humm, J.; Larson, S.; Amols, H.; Fuks, Z.; Leibel, S.; Koutcher, J. A. U.-h. w. s. c. s. a. B. W.-D.-S. a. f. a. d. f. b., Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *International Journal of Radiation Oncology, Biology, Physics* **2000**, 47, (3), 551-560.
2. Rosenman, J., Incorporating functional imaging information into radiation treatment. *Seminars in Radiation Oncology* **2001**, 11, (1), 83-92.
3. Xing, L.; Cotrutz, C.; Hunjan, S.; Boyer, A. L.; Adalsteinsson, E.; Spielman, D. M., Inverse Planning for Functional Image-Guided IMRT. *Physics in Medicine & Biology* **2002**, 47, (10), 3567-3578.
4. Alber, M.; Paulsen, F.; Eschman, S. M.; Machulla, H. J., On biologically conformal boost dose optimization. *Physics in Medicine & Biology* **2003**, 48, (10), N31-N35.
5. Yang, Y.; Xing, L., Towards biologically conformal radiation therapy (BCRT): Selective IMRT dose escalation under the guidance of spatial biology distribution. *Medical Physics* **2005**, 32, 1473-1484.
6. Pickett, B.; Woods, P.; Kurhanewicz, J.; Shinohara, K.; Roach, L. U.-. Quantification of magnetic resonance spectroscopic imaging for post permanent prostate seed implantation with 3 - 60 month follow-up. *International Journal of Radiation Oncology Biology Physics* **2001**, 51, (1), 198.
7. Zaider, M.; Zelefsky, M. J.; Lee, E. K.; Zakian, K. L.; Amols, H. I.; Dyke, J.; Cohen, G.; Hu, Y.; Endi, A. K.; Chui, C.; Koutcher, J. A., Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging. *International Journal of Radiation Oncology, Biology, Physics* **2000**, 47, (4), 1085-96.
8. Zelefsky, M. J.; Cohen, G.; Zakian, K. L.; Dyke, J.; Koutcher, J. A.; Hricak, H.; Schwartz, L.; Zaider, M., Intraoperative conformal optimization for transperineal prostate implantation using magnetic resonance spectroscopic imaging. *Cancer Journal* **2000**, 6, (4), 249-55.

-
9. DiBiase, S. J.; Hosseinzadeh, K.; Gullapalli, R. P.; Jacobs, S. C.; Naslund, M. J.; Sklar, G. N.; Alexander, R. B.; Yu, C., Magnetic resonance spectroscopic imaging-guided brachytherapy for localized prostate cancer. *Int J Radiat Oncol Biol Phys* **2002**, 52, (2), 429-38.
 10. Kim, D. H.; Margolis, D.; Xing, L.; Daniel, B.; Spielman, D., In vivo prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 T. *Magn Reson Med* **2005**, 53, (5), 1177-82.
 11. Futterer, J. J.; Scheenen, T. W.; Huisman, H. J.; Klomp, D. W.; van Dorsten, F. A.; Hulsbergen-van de Kaa, C. A.; Witjes, J. A.; Heerschap, A.; Barentsz, J. O., Initial experience of 3 tesla endorectal coil magnetic resonance imaging and ¹H-spectroscopic imaging of the prostate. *Invest Radiol* **2004**, 39, (11), 671-80.
 12. Lian, J.; Xing, L.; Hunjan, S.; Dumoulin, C.; Levin, J.; Lo, A.; Watkins, R.; Rohling, R. K.; Giaquinto, R.; Kim, D.; Spielman, S.; Daniel, B., Mapping of the Prostate in Endorectal Coil-Based MRI/MRSI and CT: a Deformable Registration and Validation Study. *Medical Physics* **2003**, 31, in press.
 13. Bookstein, F. L., Thin plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, 11, 567-585.
 14. Schreiber, E.; Xing, L., Narrow Band Deformable Registration of Prostate MRI/MRSI and CT studies. *International Journal of Radiation Oncology, Biology, Physics* **2005**, 62, 595-605.
 15. Schreiber, E.; Xing, L., Image registration with auto-mapped control volumes. *International Journal of Radiation Oncology, Biology, Physics* **2005**, 62, submitted.
 16. Yang, Y.; Xing, L., Clinical knowledge-based inverse treatment planning. *Phys Med Biol* **2004**, 49, (22), 5101-17.
 17. Yang, Y.; Xing, L., Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme. *Med Phys* **2004**, 31, (10), 2839-44.

Appendix I. manuscripts published or submitted for publication

Manuscripts:

1. Schreibmann E and Xing L: "Narrow band deformable registration of prostate magnetic resonance imaging, magnetic spectroscopic imaging, and computed tomography studies". *International Journal of Radiation Oncology, Biology, Physics* 62, 595-605, 2005.
2. Lian J, Xing L, Hunjan S, Dumoulin C, Levin J, Lo A, Watkins R., Rohling K, Giaquinto, Kim D, Spielman D, Daniel B. Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: a deformable registration and validation study. *Medical Physics* 31: 3087-3094, 2004.
3. Yang Y and Xing L: "Towards biologically conformal radiation therapy (BCRT): selective IMRT dose escalation under the guidance of spatial biology distribution ". *Medical Physics* 32, 1473-1484, 2005.
4. Yang Y, Xing L, "Inverse planning with adaptively evolving voxel-dependent penalty scheme", *Medical Physics* 31, 2839-2844, 2004.
5. Kim D, Mayer D., Xing L, Daniel B, Margolis, D., Spielman D., "In vivo prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 T", *Magnetic Resonance Imaging in Medicine*, 53, 1177-1182, 2005.
6. Schreibmann E and Xing L, "Feasibility study of beam orientation class-solutions for IMRT prostate cancer treatment". *Medical Physics* 31, 2863-2870, 2004.
7. Yang Y and Xing L: "Clinical knowledge-based inverse treatment planning". *Physics in Biology and Medicine* 49, 5101-5117, 2004.
8. Schreibmann E and Xing L: "Dose-volume based ranking of incident beam directions and its utility in facilitating IMRT beam placement". *International Journal of Radiation Oncology, Biology, Physics*, submitted.
9. Schreibmann E and Xing L: "Image registration with auto-mapped control volumes". *International Journal of Radiation Oncology, Biology, Physics*. Submitted.

Book Chapter:

Xing L, Wu Q, Yang Y and Boyer AL: Physics of IMRT and Inverse Treatment Planning, in Intensity Modulated Radiation Therapy: *A Clinical Perspective*, Mundt AF and Roeske JC. (editors), page 20-52, BC Decker Inc. Publisher, Hamilton, Canada, 2005.

PHYSICS CONTRIBUTION**NARROW BAND DEFORMABLE REGISTRATION OF PROSTATE MAGNETIC RESONANCE IMAGING, MAGNETIC RESONANCE SPECTROSCOPIC IMAGING, AND COMPUTED TOMOGRAPHY STUDIES**

EDUARD SCHREIBMANN, Ph.D., AND LEI XING, Ph.D.

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA

Purpose: Endorectal (ER) coil-based magnetic resonance imaging (MRI) and magnetic resonance spectroscopic imaging (MRSI) is often used to obtain anatomic and metabolic images of the prostate and to accurately identify and assess the intraprostatic lesions. Recent advancements in high-field (3 Tesla or above) MR techniques affords significantly enhanced signal-to-noise ratio and makes it possible to obtain high-quality MRI data. In reality, the use of rigid or inflatable endorectal probes deforms the shape of the prostate gland, and the images so obtained are not directly usable in radiation therapy planning. The purpose of this work is to apply a narrow band deformable registration model to faithfully map the acquired information from the ER-based MRI/MRSI onto treatment planning computed tomography (CT) images.

Methods and Materials: A narrow band registration, which is a hybrid method combining the advantages of pixel-based and distance-based registration techniques, was used to directly register ER-based MRI/MRSI with CT. The normalized correlation between the two input images for registration was used as the metric, and the calculation was restricted to those points contained in the narrow bands around the user-delineated structures. The narrow band method is inherently efficient because of the use of *a priori* information of the meaningful contour data. The registration was performed in two steps. First, the two input images were grossly aligned using a rigid registration. The detailed mapping was then modeled by free form deformations based on B-spline. The limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (*L-BFGS*), which is known for its superior performance in dealing with high-dimensionality problems, was implemented to optimize the metric function. The convergence behavior of the algorithm was studied by self-registering an MR image with 100 randomly initiated relative positions. To evaluate the performance of the algorithm, an MR image was intentionally distorted, and an attempt was then made to register the distorted image with the original one. The ability of the algorithm to recover the original image was assessed using a checkerboard graph. The mapping of ER-based MRI onto treatment planning CT images was carried out for two clinical cases, and the performance of the registration was evaluated.

Results: A narrow band deformable image registration algorithm has been implemented for direct registration of ER-based prostate MRI/MRSI and CT studies. The convergence of the algorithm was confirmed by starting the registration experiment from more than 100 different initial conditions. It was shown that the technique can restore an MR image from intentionally introduced deformations with an accuracy of approximately 2 mm. Application of the technique to two clinical prostate MRI/CT registrations indicated that it is capable of producing clinically sensible mapping. The whole registration procedure for a complete three-dimensional study (containing $256 \times 256 \times 64$ voxels) took less than 15 min on a standard personal computer, and the convergence was usually achieved in fewer than 100 iterations.

Conclusions: A deformable image registration procedure suitable for mapping ER-based MRI data onto planning CT images was presented. Both hypothetical tests and patient studies have indicated that the registration is reliable and provides a valuable tool to integrate the ER-based MRI/MRSI information to guide prostate radiation therapy treatment. © 2005 Elsevier Inc.

Narrow band registration model, Endorectal coil, Prostate imaging, Treatment planning.

Reprint requests to: Lei Xing, Ph.D., Stanford University School of Medicine, Department of Radiation Oncology, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Tel: (650) 498-7896; Fax: (650) 498-4015; E-mail: lei@reyes.stanford.edu

This work is supported in part by a Research Grant from the Prostate Cancer Research Program of the U.S. Department of Defense (DAMD17-03-1-0023) and the National Cancer Institute

(1 R01 CA98523-01).

Acknowledgments—The authors thank Drs. Steve L. Hancock, Christopher King, Arthur Boyer, Yong Yang, Cristian Cotrutz, Brian Thorndyke, Dong-Hyun Kim, Daniel Spielman, and Bruce Daniel for useful discussions.

Received Nov 12, 2004, and in revised form Feb 4, 2005.
Accepted for publication Feb 4, 2005.

INTRODUCTION

Endorectal (ER) coil-based magnetic resonance imaging (MRI) and magnetic resonance spectroscopic imaging (MRSI) provide high-resolution anatomic and metabolic information and allow us to better assess the extent of prostate cancer and the involvement of seminal vesicles (1–12). The new imaging information can also be used to guide the dose optimization process to selectively boost the high tumor burden regions (13–16). To use the imaging data for treatment planning, an indispensable step is the registration of ER-based MRI/MRSI and the treatment planning CT images. Presently, manual and/or automated image fusion tools implemented in most commercial treatment planning systems are based on geometric translations and rotations of the images and are generally not suitable for handling the problem because the endorectal coil displaces and distorts the prostate and surrounding tissue (17). A deformable registration procedure must be applied to map the ER-based MRI/MRSI data onto the planning CT images acquired in a normal treatment position without the insertion of ER probe (18, 19).

There have been several relevant image registration techniques reported in the literature. Zaider *et al.* (13) and Court and Dong (20) used a rigid transformation for the correction of tissue displacement. A deformable procedure based on the finite element model (FEM), in which images are described as blocks of elastic materials on which forces apply, was proposed by Bharath *et al.* (21). In this approach, the parameters that control the behavior of the elastic material and are responsible for the conversion of forces into local deformations of the elastic material are Young's elastic modulus and Poisson's ratio. Although powerful, the model has the drawback that values of the elasticity and density constant for various tissues are not readily available and have to be found by a trial and error procedure. The method also relies on using complicated software to generate a FEM mesh and masks of the involved structures. Recently, a simpler technique based on spline interpolation was proposed (18). When the first and second derivatives of the cost function can be analytically deduced, this approach reduces the problem to solving a set of system equations (22). Wu *et al.* (23) used a free-form intramodality registration between the MRI images obtained with and without inflatable ER probe acquired at the end of the study. Their study suggested that the free-form registration is stable and accurate for dealing with the problem. However, the approach is valid only for intramodality registration and entails extra MRI scans of the pelvic region with body coil. In addition, a term was introduced in their cost function to constrain the optimization to smooth deformations, posing a new problem of how to objectively select the relative weightings of different terms in the cost function.

An improvement to this method can be achieved by using a spline model with the smoothness of the deformation field assured by the interpolation between a grid of fixed control points. In this setup, the cost function is composed of a

single term and no weighting factors are required. A simple method along this line is to deduce the spline coefficients from a set of user-defined control points, as was done by Fei *et al.* (18) in a study of warping and rigid registration of MR volumes. This method was also applied to directly register prostate CT and MRI/MRSI and validated by using a series of phantom measurements (24, 25). Coselmon *et al.* (26) used a similar technique to study the accuracy of mutual-information-based CT registration of the lung at exhale and inhale respiratory states.

In general, image registration can be formulated as an optimization problem where the variables are a group of transformation parameters that lead to the best match between the input images. The match is quantified in mathematical terms by the use of a metric, which ranks a potential matching based on the image histograms, resolution, or pixel values of the involved organs. There is a vast literature on using image intensity information such as mean squares, normalized correlation, and mutual information (27) to construct metrics for guiding the registration process. The mutual information represents a popular choice when dealing with multimodality image registration. Briefly, mutual information is a criterion from information theory and is related to entropy, which is a measure of uncertainty, variability, and complexity. When each of the two images provides the most information about the other, the mutual information metric is maximal and the two images are considered to be coregistered. Methods using signed distance information from edge features have also been reported (28), in which the signed distance field of an object surface is used as the shape representation. The signed distance field is a continuous scalar function defined throughout the volume, and its value is simply the distance to the nearest point on the primary surface, with opposite signs on opposite sides of the surface. The central idea is that if the data shapes are registered correctly, the signed distance fields should match in the common coordinate system.

To register the prostate MRI/MRSI with CT images, theoretically, all we need to do is to use an optimization algorithm to minimize an mutual information metric or alike. Such setups have been applied before on the problem of CT-positron emission tomography (PET) registration (29), with splines modeling the deformations and a limited memory quasi-Newton (*L-BFGS*) algorithm (30) optimizing the system. In reality, the mutual information metric is noisy, and a deterministic optimization algorithm is generally not suitable to find the global solution. More powerful algorithms capable of statistically escaping from local minima, such as simulated annealing or genetic algorithms, are often used to attack the problem. Even with these powerful stochastic algorithms, finding the true solution of the problem can still be challenging because of the wide variety of possible pixel intensities within an organ. A narrow band formalism was devised to improve the convergence the calculation. A narrow band is composed of all points within two isosurfaces defined by the signed distance values

of $\pm d$. The method is a hybrid of the techniques based on image intensity distributions and the signed distance information (28). In essence, the technique is a two-step image registration (31) in which an organ is first represented by a data structure containing the signed distance values from objects, followed by an image registration using a pixel-based metric. A signed distance field of an object is represented through the narrow band, since the concept prevents self-intersection problems and seamlessly handles changes in topology. Previous studies have suggested that the technique improves the convergence behavior of the calculation and reduces the computational efforts (31), because sophisticated statistical considerations can be replaced with simpler pixel-based metrics computed only in the regions of clinical interest.

In this work we apply the narrow band deformable registration model to map the ER-based MRI/MRSI data onto CT images and report our implementation of the algorithm. In particular, we present our experience with the selection of model parameters, optimization algorithm, and the validation of the technique. The general reference drawn from this study is that the narrow band technique is robust and accurate for mapping information between different types of images. The approach is quite general and, with minor modifications, it should be applicable to many other deformable image registration problems in radiation therapy.

METHODS AND MATERIALS

Software platform

All calculations are implemented using an open-source software toolkit named the Insight Toolkit (ITK) (32), which consists of template-based codes for a large number of image visualization, segmentation, and registration classes. The programs contained in ITK are easily extendable, making it an ideal platform for the development of image registration methods. Concise and clear descriptions on the use of the available subroutines are provided in the ITK manual. For convenience, in the following we outline the methods used in our calculation with attention paid to the issues specific to radiation therapy image registration and to the implementation of the methods.

Registration framework

The overall image registration process of ER-based MRI/MRSI and treatment planning CT images is shown in Fig. 1. The input to

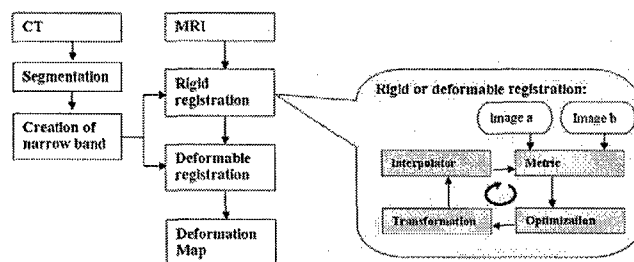


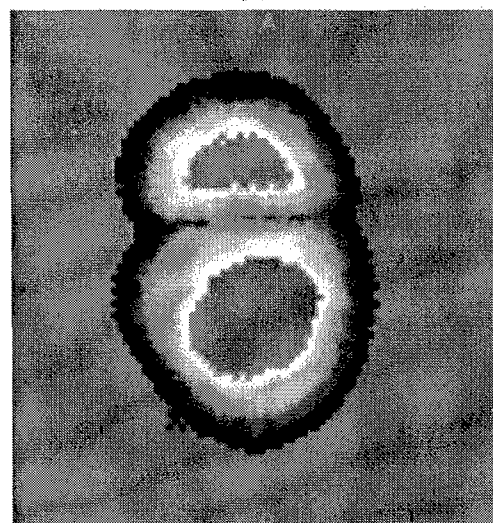
Fig. 1. Narrow band-based image registration procedure. A two-step registration is employed here to successively deal with translational and deformation miss-matches. (CT = computed tomography; MRI = magnetic resonance imaging.)



(a)



(b)



(c)

Fig. 2. (a) Magnetic resonance image zoomed in the prostate region, (b) computed tomography image of the pelvic region, and (c) the narrow band representation of rectum and prostate.

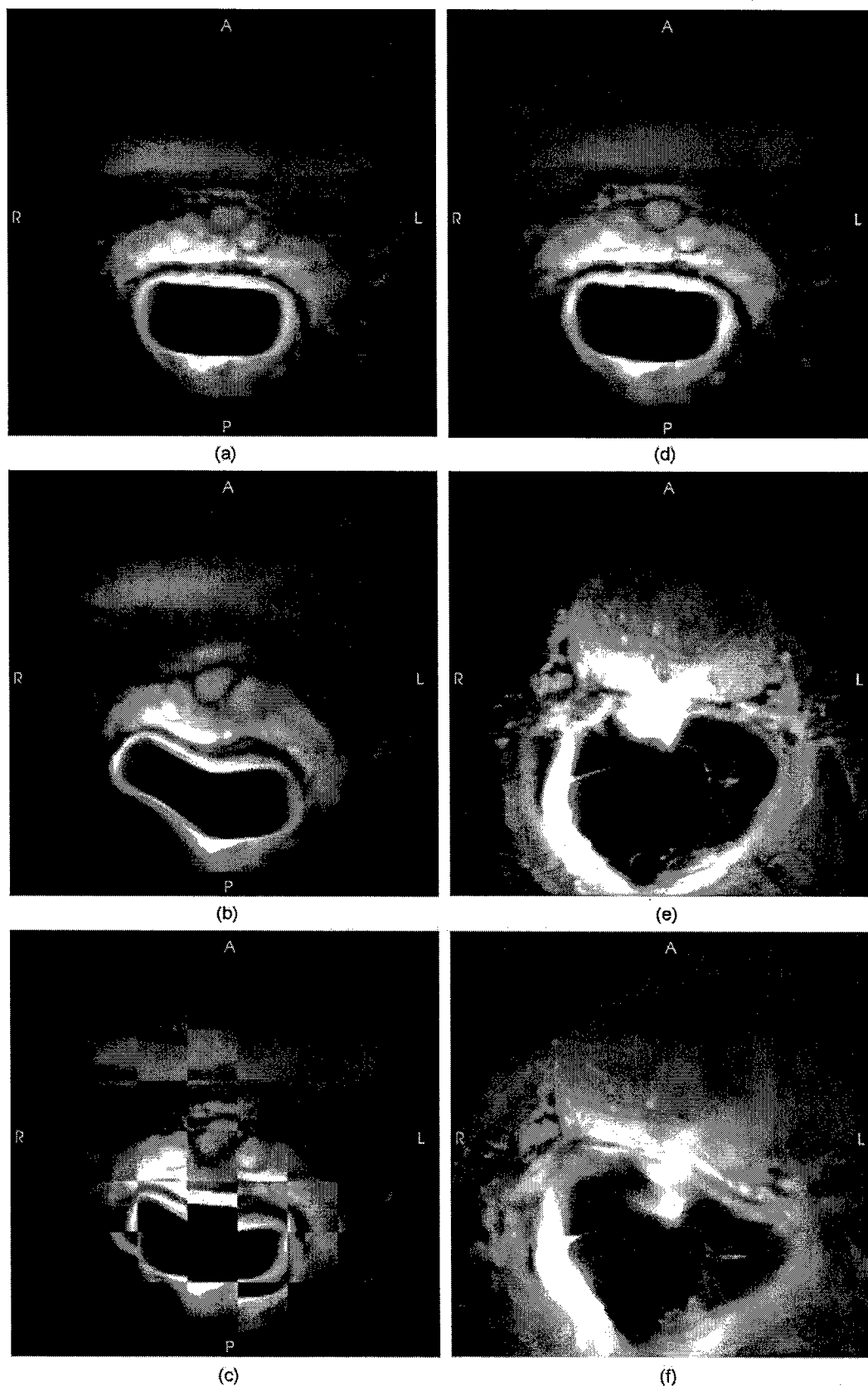


Fig. 3. Checkerboard visualization of algorithm's capability to recover from intentionally introduced deformations for two cases. The test is to register the warped image to the "gold standard" represented by the original form of the deformed magnetic resonance image. For the first case, the original and intentionally deformed images are shown in (a) and (b), respectively. The checkerboard images before and after deformable registration are shown in (c) and (d), respectively. The registration process and the results for the second case are shown in (e) to (h). Figure continues on p. 599.

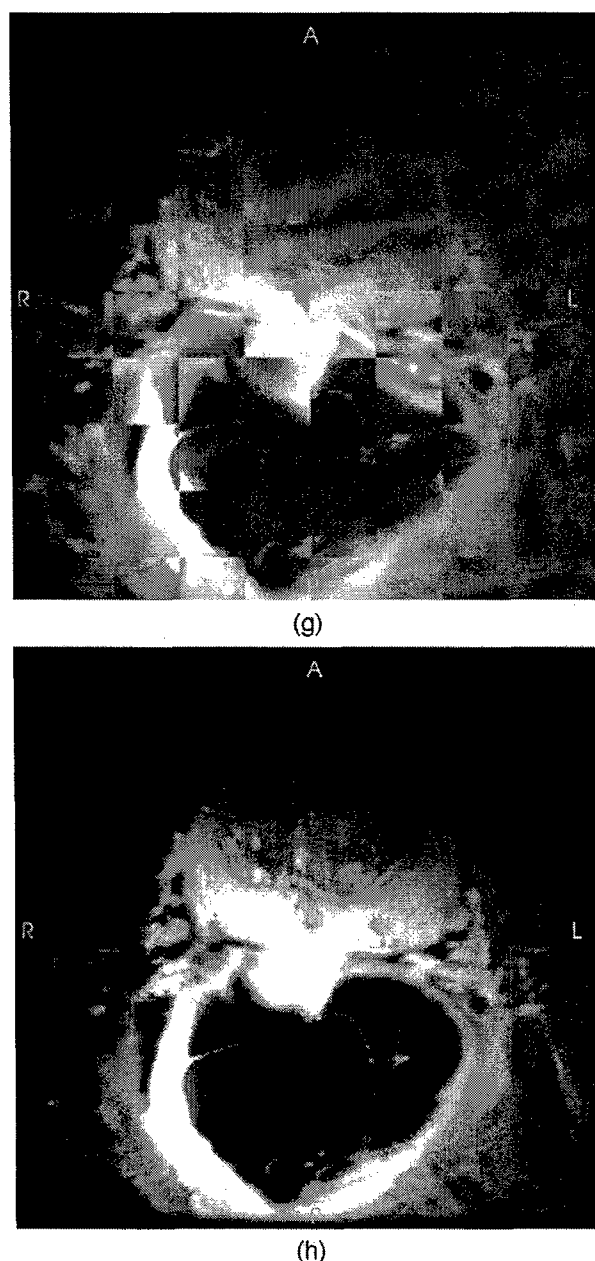


Fig. 3. (Continued)

the registration software is the images to be registered: a fixed image and a floating image, described by their intensity distributions $I_a(\mathbf{x})$ and $I_b(\mathbf{x})$, respectively. In our problem, the former one is assigned to be the treatment planning CT image and the later the ER-based MRI/MRSI. To facilitate the computational process, we separated the image registration calculation into two steps. A rigid registration is first performed to grossly align the two images and the deformable registration then follows.

For convenience, a patient is divided into a grid with N^3 cells. The corner of a lattice cell is referred to as a node and is indexed by i ($i = 1, 2, \dots, N^3$). The displacement of a node i is specified by a vector \mathbf{x}_i , and the displacement vectors, $\{\mathbf{x}_i\}$, of a collection of nodes characterize the tissue deformation. The displacement at a location \mathbf{x} on the image is deduced by fitting a polynomial expressed using the basis spline (BSpline) (33, 34) to the grid nodes \mathbf{x}_i . Unlike other spline models, the BSplines are locally controlled.

That is, the displacement of an interpolation point is influenced only by that of the closest grid points, and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations.

A signed narrow band is defined on the CT image for prostate and rectum based on the physician-delineated contour. A narrow band is composed of a number of nodes with known distances to the contours, which introduces additional positional information in the representation. Typical narrow bands for rectum and prostate are shown in Fig. 2c as a gray-level image. Pixel intensities correspond to the signed distances with dark/white values representing the negative/positive distances. The zero level set is represented with a medium gray intensity. The gray background in this image is not part of the narrow band structure.

The task of standard image registration is to find the transformation matrix, $T(\mathbf{x})$, that maps an arbitrary point \mathbf{x} on the fixed image to the corresponding point \mathbf{x}' on the floating image (or vice versa) so that the best possible match, as measured by the registration metric, is achieved. The choice of the metric for ranking different possible matching differentiates one algorithm from the others and is fundamental to the success of the image registration. In our calculation, each organ in the floating image is represented by the data structure of the signed distance or by the narrow band surrounding the edge of the organ. The narrow band acts as a shape representation model of an anatomic structure. It is the fixed or target object in the registration process so that the band does not need to be regenerated for each evaluation of the metric (31). As shown in Fig. 1, the image registration proceeds in an iterative fashion. The matrix coefficients of $T(\mathbf{x})$ are the node displacements and are adjusted iteratively to minimize the normalized correlation between the two images defined as:

$$f = \frac{\sum_{i=1}^N I_a(\mathbf{x}_i) I_b(T\mathbf{x}_i)}{\sqrt{\sum_{i=1}^N I_a^2(\mathbf{x}_i) \sum_{j=1}^N I_b^2(T\mathbf{x}_j)}} \quad (1)$$

where i and j are the node indices within the narrow band on the fixed image, $I_a(\mathbf{x}_i)$ is the intensity of the node at \mathbf{x}_i on the fixed image a , and $I_b(T\mathbf{x}_i)$ is the intensity of the image b at where the node \mathbf{x}_i is mapped. The metric uses both voxel intensity and delineated structures as complementary information for the registration.

The *L-BFGS* algorithm for optimization

Optimization of the normalized correlation function with respect to the displacements of the nodes, $\{\mathbf{x}_i\}$, yields the transformation coefficients $T(\mathbf{x})$ that map the points on image a to image b . Since the two images do not necessarily have the same size, an interpolation may be needed to compute intensity at a mapped point $\mathbf{x} = T(\mathbf{x})\mathbf{x}$. To facilitate the optimization it is preferable that both the deformable model and the metric are differentiable (29). This condition is satisfied for the system that we are dealing with, as demonstrated in a previous mathematical work (35).

We used the limited memory *BFGS* algorithm *L-BFGS* (30), which is known for its superior performance in dealing with high-dimensionality problems, to optimize the system here. *L-BFGS*, as compared with a conventional gradient search method such as the Newton's algorithm, does not require the exact inverse Hessian matrix. For an N -dimensional problem, only $4mN$ operations are needed in *L-BFGS* to calculate the descent direction at an

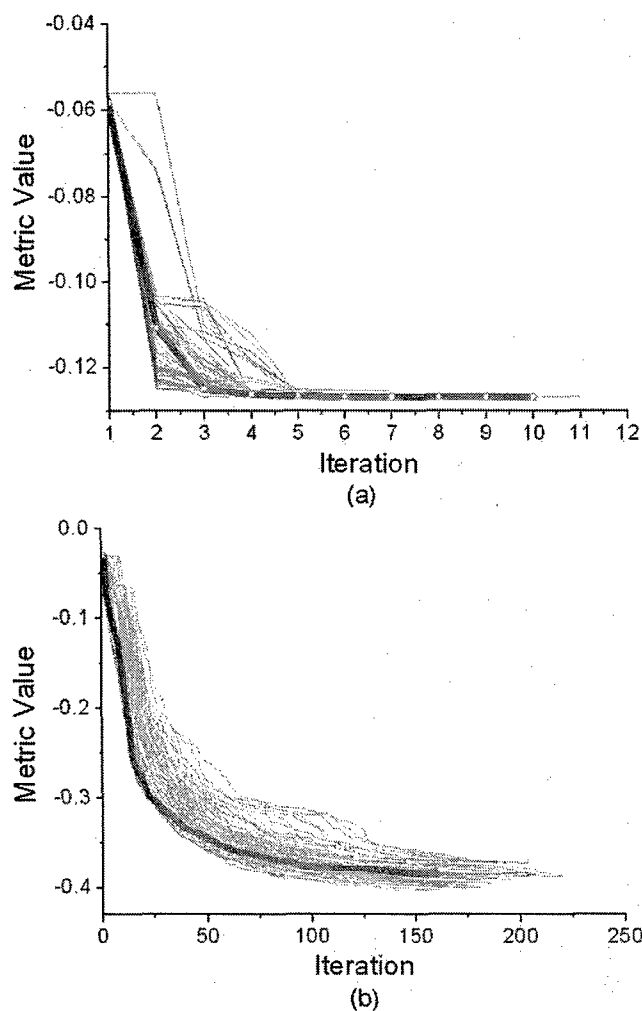


Fig. 4. Narrow band metric as a function of iteration step for calculations with 100 different initial values. Both rigid (a) and deformable (b) registration calculation are plotted. In both cases, a typical convergence is displayed in red. Rigid registration convergence is achieved after less than 10 iterations. Deformable registration converges in approximately 100 iterations.

iteration k , indirectly from the m previous values of $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ of the system variable \mathbf{x} (30, 36). A value of $m = 5$ is recommended. For a large N system, the number of operations becomes significantly less than $O(N^2)$, as involved in *BFGS* calculation. Starting from a positive definitive approximation of the inverse Hessian \mathbf{H}_0 at \mathbf{x}_0 , *L-BFGS* derives the optimization variables by iteratively searching through the solution space. At an iteration k , the calculation proceeds as follows:

- Determine the descent direction $\mathbf{p}_k = -\mathbf{H}_k \nabla f(\mathbf{x}_k)$;
- Line search with a step size $\alpha_k = \arg \min_{\alpha \geq 0} f(\mathbf{x}_k + \alpha \mathbf{p}_k)$, where $\alpha \geq 0$ is the step size defined in the *L-BFGS* software package;
- Update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$;
- Compute \mathbf{H}_{k+1} with the updated \mathbf{H}_k .

At each iteration a backtracking line search is used in *L-BFGS* (30) to determine the step size of movement to reach the minimum of f along the ray $\mathbf{x}_k + \alpha \mathbf{p}_k$. For convergence α has to be chosen such that a sufficient decrease criterion is satisfied, which depends on

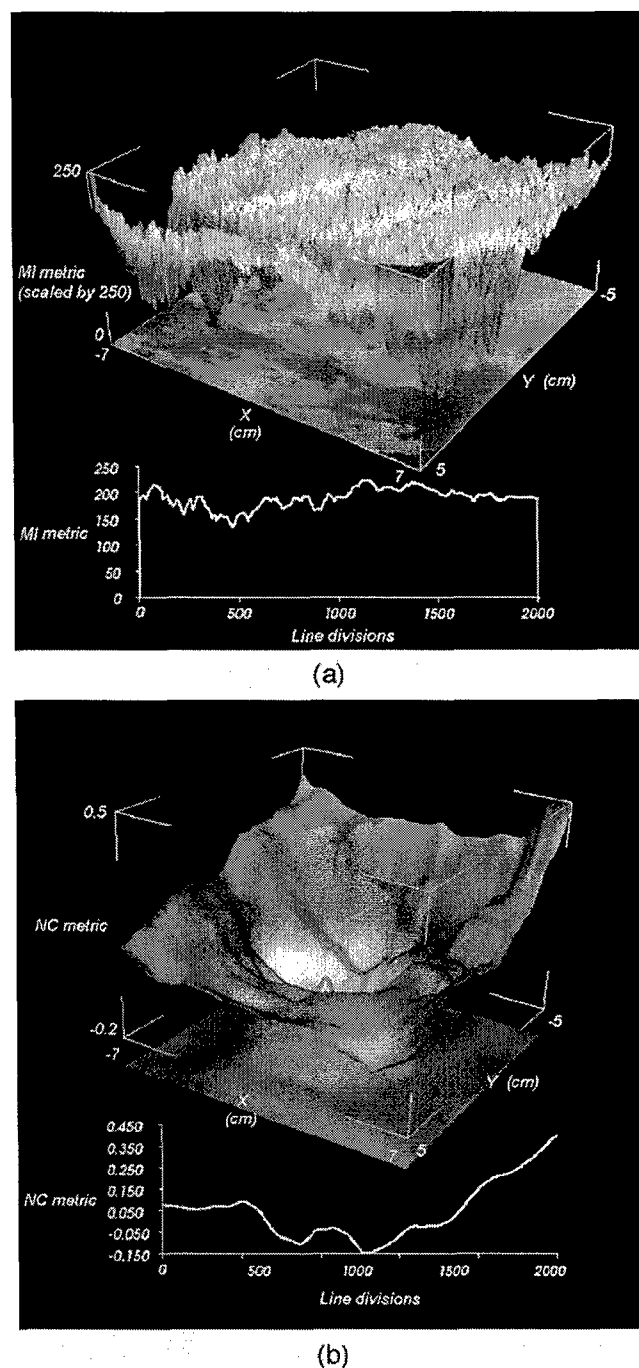
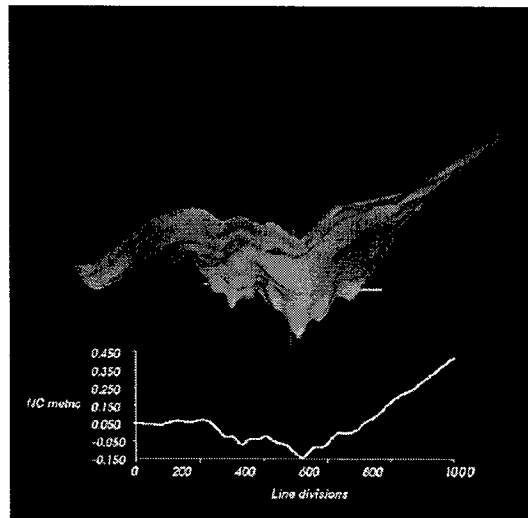


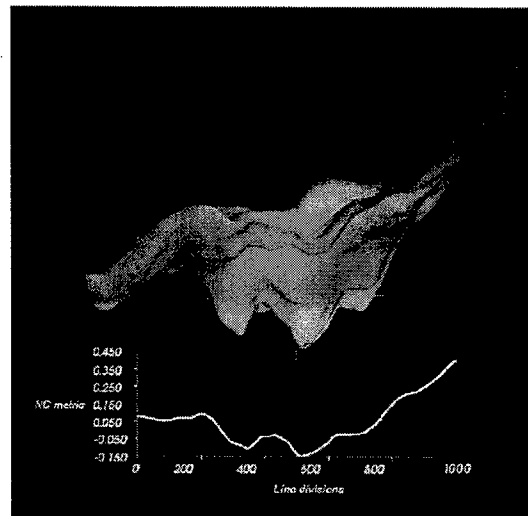
Fig. 5. The mutual information metric (a) and the narrow band metric (b) as a function of relative displacement in the xy plane. Two identical magnetic resonance images (from Fig. 2a) were used as the input fixed and floating images. The insert in each panel shows the corresponding metric when the displacement is confined to the x -direction. The mutual information metric is seen to be very "noisy" and multiple maxima exist. The narrow band metric, on the other hand, is much smoother and has a single minimum.

the local gradient and function value and is specified in *L-BFGS* by the Wolfe conditions (30).

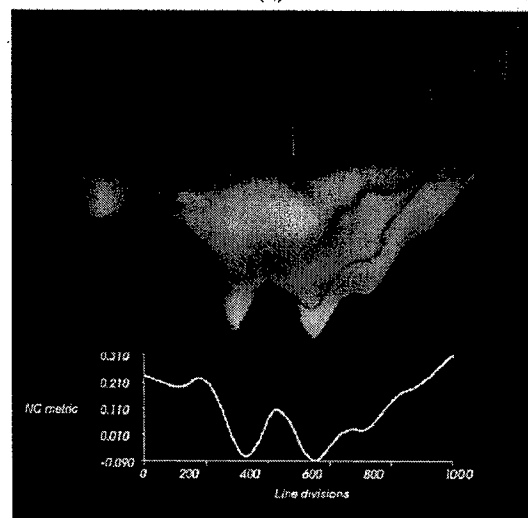
During an image registration process, the above iterative calculation based on *L-BFGS* algorithm continues until either a pre-set maximum number of iterations (typically, 500 iterations) is reached or the following stopping criterion is fulfilled:



(a)



(b)



(c)

$$\frac{\|\nabla f(\mathbf{x}_k)\|_2}{\max(1, \|\mathbf{x}_k\|_2)} < \varepsilon \quad (2)$$

In our calculations, we chose $\varepsilon = 10^{-6}$. It is possible that *L-BFGS* may produce unrealistic deformations in finding the minimum of the metric. Similar to previous investigators (29), we used the bounded version of *L-BFGSB* (37), where the variables representing deformation vectors are restricted to within certain limits. The convergence behavior of the *L-BFGS* algorithm is studied by starting the registration calculation with different random initial positions. Intermediate values obtained during the optimization calculations are recorded and plotted for comparison.

Image acquisition

The MRI images (marked as FSE-xT2 fat) were acquired using a 3-Tesla MR scanner (Signa; GE Medical Systems, Milwaukee, WI). The radiofrequency (RF) excitation was achieved by using the whole body birdcage resonator, and the MR signal was received using a 4-element phased-array antenna (GE Medical Systems) combined with a rigid single loop receiver-only surface coil with a fixed geometry that enables optimal tuning and matching for use at 3T. The coil dimensions are similar to transrectal ultrasound transducers used for routine sonographically guided prostate imaging and biopsy. Patient CT images were acquired using a PQ5000 CT Scanner (Philips Medical Systems, Cleveland, OH).

Behavior of the metric function

Monitoring the change of the metric function under the relative displacement of two input images provides a useful examination of the behavior of the solution space for image registration. To better understand the narrow band based normalized correlation metric, we used two identical images as input and computed the function by successively translating them along the x- and y-directions. In this case, it is intuitively conceivable that the value of the metric function is minimum when they are perfectly aligned and increases as the two images are displaced away from each other. A desirable metric function should exhibit a single distinct global minimum and vary smoothly as the two images slide away from each other. For comparison, the same calculation was also performed for conventional mutual information metric (27). The two types of functions are presented and reviewed by using 3D graphs, where surface height represents the metric value.

Registration accuracy

In this study, the input images involved an MRI as the fixed image and the same image with intentionally introduced deformations as the floating images. The ability of the narrow band algorithm to restore from the deformation was tested. The original image in this study serves as the "gold standard" for the evaluation.

Fig. 6. The narrow band metric as a function of relative displacement in the xy plane and in the x-direction when the narrow band width takes a value of 3 mm (a), 5 mm (b), and 15 mm (c). For a small width, the narrow band has a small capture radius with the minima of the metric accentuated. Increasingly narrow band width enlarges the capture radius. The computation becomes more demanding in this case owing to the increased number of narrow band nodes. A width of approximately 6 mm seems to balance the two limiting factors.

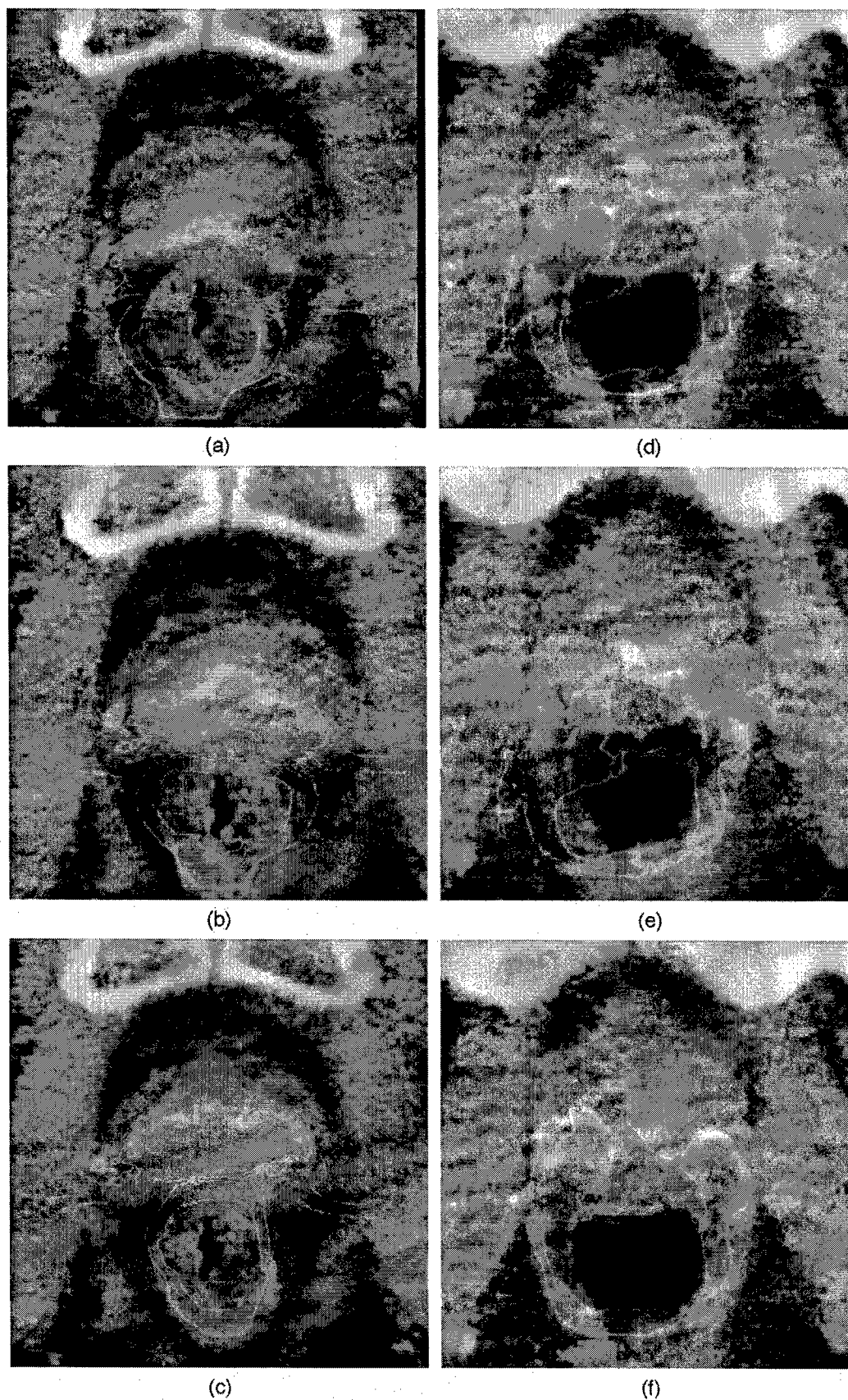


Fig. 7. Registration result for two clinical cases. The first row shows the colored overlay of the endorectal-based magnetic resonance and computed tomography images before and after rigid and after deformable image registration for the first case. The computed tomography image is used as the background. The second row shows the same for the second clinical case.

Figures 3a and 3b show the original and deformed MR images for the first patient. The deformation was generated by assigning the nodes in the original image with a set of random displacement vectors. Eight grid points with spacing of 2.4 cm were used in the calculation. We used the proposed algorithm to restore the distortions and evaluated the algorithm's ability to recover the gold standard. The difference between the original and restored images was quantified in terms of displacements of visible anatomic landmarks. The images before and after restoration were reviewed by using the popular checkerboard display tool, where images to be compared are merged together in a chess-like pattern. The original and deformed images for the second case are shown in Figs. 3e and 3f. Similar calculation and analysis were performed for this case.

Case study

The narrow band based image registration algorithm was applied to register ER-based MRI and CT prostate images for two clinical cases. The MRI and CT images were acquired using the protocols described earlier. The colored overlay of the two types of images was generated before and after the narrow band image registration calculations for assessment of the results. For each case, the convergence of the calculation was also studied.

RESULTS AND DISCUSSION

Convergence analysis

The algorithm's ability to obtain the same result with different starting conditions was examined. In Fig. 4, we plot the metric value as a function of iteration step starting from 100 different initial transform parameters, $\{x_i\}$, for the first case shown in Fig. 3. The input images in this study included an ER-based MRI as the floating image and the treatment planning CT as the fixed image. Two types of registrations were studied here: a rigid (Fig. 4a) and a deformable (Fig. 4b) registration. For the rigid registration, the normalized correlation converges to -0.130 , after 11 iterations. The final value of the metric is found to be independent of the initial starting conditions. For deformable registration, the minimum metric value was found to be -0.399 . The longest run in obtaining this solution was 218 iterations. Once again, the result was independent of the initial starting conditions. For each type of registration, a typical evolution of the metric as a function of the iteration step is displayed in Fig. 4 in the red curve.

A narrow band is a compact representation of a structure because only points adjacent to its border are used. This permits a reduction of memory requirements for the optimization process, with computation times reduced by 1 or 2 orders of magnitude. For this particular study, *L-BFGS* needed approximatively 100 iterations to converge. Compared with some other potentially more powerful evolutionary algorithms, an analysis not shown here revealed that a typical $1 + 1$ optimizer (38) needs as much as 100,000 iterations for convergence. This is reflected in the computation times, where 15 min are needed by *L-BFGS* for a full study of $256 \times 256 \times 30$ voxels, vs. more than 1 h needed by a $1 + 1$ optimizer.

Behavior of mutual information and narrow band metric functions

In Fig. 5, we plot the mutual information and narrow band metric values when the two identical input images are translated on the x-y plane. Because the attempt here is to register an MRI image (Fig. 5a) with itself, the position of global minimum is known to correspond to a null translation. The variation of the metrics as a function of the translation in the x-direction is plotted as inset in Fig. 5. The mutual information metric space, presented in Fig. 5a, was obtained with settings of 100 spatial samples and 0.4 standard deviation (27). While the general trend toward a global solution located at the center can be deduced, the metric space is noisy and many local minima are present. This metric is thus not suitable to be used with gradient-based optimizers.

As can be seen from Fig. 5b, the search space corresponding to the normalized correlation metric is much smoother than that of the mutual information metric (27) and has a distinct minimum at the expected location. The local minimum at the left is resulted from the alignment of the rectum with the prostate. However, this minimum is shallower, indicating a poor fit, owing to the difference in rectum/prostate shapes. We note that with the selection of different narrow band widths it is possible to change the capture radius of the metric function (Fig. 6). As the narrow band width increases, the metric capture radius increases; and at the same time, the computation time needed to find the solution also increases since more nodes are involved. In this trade-off, we found a narrow band width of approximately 6 mm to be optimal for practical application.

The narrow band metric uses both voxel intensity and delineated structures information for the registration. For prostate MRI-CT registration, an intensity-based metric such as the mutual information is intractable owing to the lack of clearly differentiable anatomic structures in the CT images (Fig. 2b) and the presence of a broad range of voxel intensities in the MRI image (Fig. 2a). The narrow band metric compensates for the inconveniences by restricting the calculation to selected regions of interest. The narrow band not only represents the shape of the contour but also a transition zone around it. This results in a large capture radius in the optimization calculation. As compared with control-points based methods, where the registration accuracy depends heavily on the specification of the control points, the narrow band based image registration entails little information from the user other than the delineation of the involved organs, which one has to do anyway for the purpose of radiation treatment planning.

Evaluation of the algorithm's ability of restoring from an intentionally introduced deformation

For the deformation depicted in Fig. 3b, ideally, the application of a deformable image registration technique should be able to restore its original shape when the undeformed image (Fig. 3a) is used as a reference. To better visualize the intentionally introduced deformation, in Fig.

3c we show the checkerboard image of the original and distorted image. The initial deformations up to 10 mm in the original image are clearly visible in the checkerboard transition zone. To use the narrow band based registration tool, the organs in the reference image were outlined and narrow bands were generated. The restored image is shown in Fig. 3d along with the original reference image in a checkerboard fashion. Recovered and gold standard reference images differ by less than 2 mm, as assessed by comparing displacements of the anatomic landmarks between neighborhood squares of the checkerboard. The original and restored images are highly similar, with virtually no difference visible in the checkerboard. Similar level of accuracy was found for the second case studied (Figs. 3e–3h). In both cases, we found that the calculation converged to the anticipated solution with fewer than 220 iterations.

Case study

In Fig. 7, we show the prostate MRI-CT registration results for two clinical cases. The same colored overlay is constructed before (Fig. 7a for Case 1 and Fig. 7d for Case 2) and after registration calculation (Fig. 7c for Case 1 and Fig. 7f for Case 2). In the colored overlay, the original CT image is used as the image background. The colored overlay represents the MRI image, with a colored scheme corresponding to different MRI pixel intensities, red for high intensities and blue for low intensities. Because of the use of the ER probe, the shapes of both rectum and prostate were deformed. As can be seen from Figs. 7a or 7d, the anterior surface of the rectum was flattened. After the rigid registration calculation, we noticed that the gross misalignment was removed (Figs. 7b and 7e). As expected, there was still substantial deformation after the initial rigid registration that needs to be corrected by a deformable image registration algorithm. In the next stage of the registration calculation, the deformable model was “switched on,” which restored the shapes of the rectum and prostate to their regular forms as observed in the CT images. This can be seen from Figs. 7c for the first case or Fig. 7f for the second case.

In our calculation, we observed that the regular grid of BSpline control points could be mapped to a region outside the narrow band. While it seems that this does not directly affect the accuracy of the method, it may prolong the calculation by computing the displacements in regions where no metric information is available. Setups have been proposed to adapt the splines control mesh to regions where deformation is found to be significant (39), and the extension of the method would allow us to use the BSpline control points defined only in the regions within the narrow band. Implementation of this type of technique should further reduce the computation time required to find the optimal solution.

CONCLUSIONS

We have applied the narrow band algorithm for the registration of ER-based prostate MRI and treatment planning CT images. The narrow band is a compact representation of a structure because only pixels close to the structure boundaries are considered. A distance map around structures permits the use of simple registration metrics without relying on introducing control points. Both validation and patient studies have indicated that the registration is reliable and provides a valuable tool for integrating the ER-based MRI information into radiation therapy treatment planning. Because the contours of the organs are needed anyway for treatment planning purposes, the approach involves little additional work in constructing the narrow bands. Our study indicated that the narrow band metric has a smooth metric space, which permits us to use deterministic algorithms such as *L-BFGS* to optimize the system. The narrow band registration procedure was found to be fast, stable, and capable of providing practically acceptable accuracy. Finally, the method works with CT images of any quality, as important information is contained within the narrow band representation specified by the user. This might have a significant practical implication with the prevalence of KeV and MeV cone-beam CT in clinical practice.

REFERENCES

1. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24–0.7-cm³) spatial resolution. *Radiology* 1996;198:795–805.
2. Scheidler J, Hricak H, Vigneron DB, et al. Prostate cancer: Localization with three-dimensional proton MR spectroscopic imaging—clinicopathologic study. *Radiology* 1999;213:473–480.
3. Pickett B, Kurhanewicz J, Coakley F, Shinohara K, Fein B, Roach M. Use of MRI and spectroscopy in evaluation of external beam radiotherapy for prostate cancer. *Int J Radiat Oncol Biol Phys* 2004;60:1047–1055.
4. Schnall MD, Imai Y, Tomaszewski J, Pollack HM, Lenkinski RE, Kressel HY. Prostate cancer: Local staging with endorectal surface coil MR imaging. *Radiology* 1991;178:797–802.
5. Schnall MD, Lenkinski RE, Pollack HM, Imai Y, Kressel HY. Prostate: MR imaging with an endorectal surface coil. *Radiology* 1989;172:570–574.
6. Presti JC, Hricak H, Narayan PA, Shinohara K, White S, Carroll PR. Local staging of prostatic carcinoma: Comparison of transrectal sonography and endorectal MR imaging. *AJR Am J Roentgenol* 1996;166:103–108.
7. Hricak H, Dooks GC, Jeffrey RB, et al. Prostatic carcinoma: Staging by clinical assessment, CT, and MR imaging. *Radiology* 1987;162:331–336.
8. Perrotti M, Kaufman RP Jr, Jennings TA, et al. Endo-rectal coil magnetic resonance imaging in clinically localized prostate cancer: Is it accurate? *J Urol* 1996;156:106–109.
9. Futterer JJ, Scheenen TW, Huisman HJ, et al. Initial experience of 3 Tesla endorectal coil magnetic resonance imaging and 1H-spectroscopic imaging of the prostate. *Invest Radiol* 2004;39:671–680.
10. Jager GJ, Severens JL, Thornbury JR, de La Rosette JJ, Ruijs

- SH, Barentsz JO. Prostate cancer staging: Should MR imaging be used? A decision analytic approach. *Radiology* 2000;215:445–451.
11. Quinn SF, Franzini DA, Demlow TA, *et al.* MR imaging of prostate cancer with an endorectal surface coil technique: Correlation with whole-mount specimens. *Radiology* 1994; 190:323–327.
12. Kim DH, Margolis D, Xing L, Daniel B, Spielman D. *In vivo* prostate magnetic resonance spectroscopic imaging using two-dimensional J-resolved PRESS at 3 Tesla. *Magn Reson Med* 2005; in press.
13. Zaider M, Zelefsky MJ, Lee EK, *et al.* Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging. *Int J Radiat Oncol Biol Phys* 2000;47:1085–1096.
14. Xing L, Cotrutz C, Hunjan S, Boyer AL, Adalsteinsson E, Spielman D. Inverse planning for functional image-guided intensity-modulated radiation therapy. *Phys Med Biol* 2002; 47:3567–3678.
15. Das SK, Miften MM, Zhou S, *et al.* Feasibility of optimizing the dose distribution in lung tumors using fluorine-18-fluorodeoxyglucose positron emission tomography and single photon emission computed tomography guided dose prescriptions. *Med Phys* 2004;31:1452–1461.
16. Ling CC, Humm J, Larson S, *et al.* Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys* 2000;47:551–560.
17. Hirose M, Bharatha A, Hata N, *et al.* Quantitative MR imaging assessment of prostate gland deformation before and during MR imaging-guided brachytherapy. *Acad Radiol* 2002;9: 906–912.
18. Fei B, Kemper C, Wilson DL. A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes. *Comput Med Imag Graph* 2003;4:267–281.
19. Christensen GE, Carlson B, Chao KS, *et al.* Image-based dose planning of intracavitary brachytherapy: Registration of serial-imaging studies using deformable anatomic templates. *Int J Radiat Oncol Biol Phys* 2001;51:227–243.
20. Court LE, Dong L. Automatic registration of the prostate for computed-tomography-guided radiotherapy. *Med Phys* 2003; 30:2750–2757.
21. Bharath A, Hirose M, Hata N, *et al.* Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. *Med Phys* 2001;28: 2551–2560.
22. Lu W, Chen ML, Olivera GH, Ruchala KJ, Mackie TR. Fast free-form deformable registration via calculus of variations. *Phys Med Biol* 2004;49:3067–3087.
23. Wu X, Dibiase SJ, Gullapalli R, Yu CX. Deformable image registration for the use of magnetic resonance spectroscopy in prostate treatment planning. *Int J Radiat Oncol Biol Phys* 2004;58:1577–1583.
24. Lian J, Hunjan S, Dumoulin C, *et al.* Integrating deformable MRI/MRSI and CT image registration into the prostate IMRT treatment planning [Abstract]. *Int J Radiat Oncol Biol Phys* 2003;57(Suppl. 2):S207.
25. Lian J, Xing L, Hunjan S, Spielman B, Daniel B. Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study. *Med Phys* 2004; 31:3087–3094.
26. Coselmon MM, Balter JM, McShan DL, Kessler ML. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med Phys* 2004;31:2942–2948.
27. Viola P, Wells WM. Alignment by maximization of mutual information. *Int J Comput Vision* 1997;24:137–154.
28. Borgefors G. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Transact Pattern Anal Machine Intel* 1988;10:849–865.
29. Mattes D, Haynor RD, Vesselle H, Lewellen KT, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Transact Med Imag* 2003;22:120–128.
30. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Mathemat Progr* 1989;45:503–528.
31. Ng L, Ibanez L. Narrow band to image registration in the Insight Toolkit. *Lecture Notes Comput Sci* 2003;2717:271–280.
32. Ibanez L, Schroeder W, Ng L, Cates J. ITK software guide. Clifton Park, NY: Kitware; 2003.
33. Lee S, Wolberg G, Chwa KY, Shin SY. Image metamorphosis with scattered feature constraints. *IEEE Transact Visualiz Comput Graph* 1996;2:337–354.
34. Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. *IEEE Transact Visualiz Comput Graph* 1997;3:228–244.
35. Thevenaz P, Unser T. Spline pyramids for inter-modal image registration using mutual information. *IEEE Transact Med Imag* 1997;16:187–198.
36. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C. 2nd ed. Cambridge, England: Cambridge University Press; 1992.
37. Byrd RH, Lu P, Nocedal J, Zhu CY. A limited memory algorithm for bound constrained optimization. *SIAM J Numer Anal* 1995;5:1190–1208.
38. Wienholt W. Entwurf neuronaler Netze: Ein daten-basiertes Verfahren zur Modellierung nichtlinearer Systeme mit Hilfe evolutionärer Algorithmen und unscharfer Logik. Frankfurt am Main: Harri Deutsch; 1996.
39. Camara O, Colliot O, Delso G, Bloch I. 3D Nonlinear PET-CT image registration algorithm with constrained free-form deformations. *Proc 3rd IASTED Int Conf Visualiz Imaging Image Process* 2003;1:516–521.

Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study

J. Lian,^{a)} L. Xing,^{b)} and S. Hunjan

Department of Radiation Oncology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, California 94305

C. Dumoulin

GE Corporate Research and Development Center, Schenectady, New York 12309

J. Levin

Department of Radiology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, California 94305

A. Lo

Department of Radiation Oncology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, California 94305

R. Watkins, K. Rohling, and R. Giaquinto

GE Corporate Research and Development Center, Schenectady, New York 12309

D. Kim, D. Spielman, and B. Daniel

Department of Radiology, Stanford University School of Medicine, 875 Blake Wilbur Drive, Stanford, California 94305

(Received 1 June 2004; revised 26 July 2004; accepted for publication 24 August 2004; published 28 October 2004)

The endorectal coil is being increasingly used in magnetic resonance imaging (MRI) and MR spectroscopic imaging (MRSI) to obtain anatomic and metabolic images of the prostate with high signal-to-noise ratio (SNR). In practice, however, the use of endorectal probe inevitably distorts the prostate and other soft tissue organs, making the analysis and the use of the acquired image data in treatment planning difficult. The purpose of this work is to develop a deformable image registration algorithm to map the MRI/MRSI information obtained using an endorectal probe onto CT images and to verify the accuracy of the registration by phantom and patient studies. A mapping procedure involved using a thin plate spline (TPS) transformation was implemented to establish voxel-to-voxel correspondence between a *reference* image and a *floating* image with deformation. An elastic phantom with a number of implanted fiducial markers was designed for the validation of the quality of the registration. Radiographic images of the phantom were obtained before and after a series of intentionally introduced distortions. After mapping the distorted phantom to the original one, the displacements of the implanted markers were measured with respect to their ideal positions and the mean error was calculated. In patient studies, CT images of three prostate patients were acquired, followed by 3 Tesla (3 T) MR images with a rigid endorectal coil. Registration quality was estimated by the centroid position displacement and image coincidence index (CI). Phantom and patient studies show that TPS-based registration has achieved significantly higher accuracy than the previously reported method based on a rigid-body transformation and scaling. The technique should be useful to map the MR spectroscopic dataset acquired with ER probe onto the treatment planning CT dataset to guide radiotherapy planning. © 2004 American Association of Physicists in Medicine. [DOI: 10.1118/1.1806292]

I. INTRODUCTION

The introduction of endorectal (ER) surface coils significantly improves the spatial resolution and signal-to-noise ratio (SNR) of prostate MR and MR spectroscopic imaging.¹⁻⁸ The new MRI/MRSI tool provides an unprecedented means for us to characterize the location(s) and volume(s) of intraprostatic lesion(s) and to evaluate the possible capsular penetration, invasion of neurovascular bundle, and seminal vesicle involvement.^{2-4,9-13} The information derived from the new imaging modality is also valuable for guiding radiation treatment planning to escalate radiation doses according to the regional tumor burden.¹⁴⁻¹⁸ In practice, the use of an

ER coil severely distorts the prostate and surrounding organs. On the other hand, current radiation treatment planning is performed on the CT images without distortion. CT has high geometric accuracy and provides valuable electron density information needed for accurate dose calculation. In order to use ER-based image data to guide radiation treatment planning, it is imperative to develop a method to map the information in the ER-based MRI/MRSI to the corresponding location in CT images.^{19,20} Zaider *et al.*¹⁷ and Mizowaki *et al.*²¹ have reported a translation and scaling based registration method to map MRS positive volumes onto the CT and ultrasound images. In their approach, the coordinates of

the boundary and the center of mass were used to linearly interpolate the positions of the mapped voxels. Although a clinically acceptable mean difference (2.4 mm) between the predicted and measured positions was reported, larger discrepancy was found for regions with more severe distortion (≥ 4 mm).

In order to fully use the functional data to guide radiation treatment planning, a mapping method with computation efficiency and acceptable accuracy is needed. The purpose of this paper is to present a thin plate spline (TPS)-based deformable registration to improve the previously reported nondeformable MRS and CT mapping technique and test the registration accuracy using a series of phantom measurements. The TPS transformation is a well-established mathematical method and its central idea is to find a continuous transformation to minimize the difference between the control points in two images. Since its first introduction into medical image analysis,²⁶ the TPS has been successfully used on several applications. A two step registration scheme (rigid body registration and TPS warping) was employed to make comparisons of MR images in interventional MRI guided radiofrequency ablation to determine whether a tumor is adequately treated.^{22,23} In order to map changes in the shape and position of the liver between inhale and exhale breath held CT models of a patient, a mutual information (MI) based alignment with TPS warping was proposed.²⁴ A TPS transformation based technique has also been found useful to correct image distortion in fluoroscopic images.²⁵ We believe this method is a good tradeoff between computation complexity and registration accuracy and should be well suited in mapping deformed voxels of MRS onto CT image.

II. METHODS AND MATERIALS

A. Phantom construction and imaging

Tissue equivalent bolus material was used to construct the 2D phantoms which simulate the axial sections of the patient dataset. The bolus, made of vinyl gel, is elastic and has a density close to that of water. Ten to fifteen metal fiducial landmarks were embedded into each phantom. The phantoms were held in a custom made plastic holder, allowing them to be constricted and deformed in specifically chosen regions (Fig. 1). The radiographic images were then acquired in anterior-posterior (AP) and lateral (LT) direction using a Ximatron Radiotherapy Simulator (Varian Medical Systems, Palo Alto, CA). An analysis of the AP/LT images for each phantom revealed the geometric locations of the fiducial markers before and after the distortion.

B. Patient image acquisition

Patient MRI was acquired on a 3-Tesla MR scanner (Signa; GE Medical Systems, Milwaukee, WI). RF excitation was achieved by using the whole body birdcage resonator, and the MR signal was received using a 4-element phased-array antenna (G.E. Medical Systems, Milwaukee, WI) combined with a rigid single loop receiver-only surface coil with

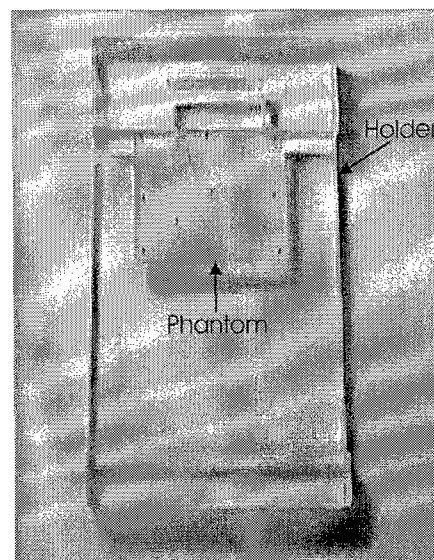


FIG. 1. A photo of the deformable phantom with implanted landmarks in a holder. The landmarks are used to calculate the registration error.

a fixed geometry that enables optimal tuning and matching for use at 3 T. The coil dimensions are similar to transrectal ultrasound transducers used for routine sonographically-guided prostate imaging and biopsy. The ER-induced distortion of MRSI is very close to that of MRI. We show MR images in this study because they have higher image quality than MRSI. MR images were acquired using axial fast spin echo TR/TE: 6000/80.5 ms, echo train length: 48 ms, FOV: 10 cm, Matrix: 512×256 and Resolution: $195 \times 390 \mu\text{m}$. Patient CT images were acquired using a PQ5000 CT Scanner (Philips Medical Systems, Cleveland, OH). Three patients were recruited for the scanning. They have stage T1/T2 disease and pretreatment prostate specific antigen (PSA) level was ≤ 10 ng/ml.

C. Image mapping method

After the acquisition of CT and MRS/MR image, the prostate volumes were contoured by an experienced oncologist. The rotation operator was applied to adjust the relative tilt between two volumes. The axial slices of CT and MR data set were resampled using 1 mm interval. We aligned CT and MR slices with reference to the apex and base of the glands. Four to eight control points were placed in each pair of slices. The control points were only put along the contour of the gland and they are featured points such as corners and intersections of edges. Lastly the TPS transformation was applied on each pair of slices to establish a mapping relationship between voxels of MRI and CT. For convenience, henceforth, the nondistorted CT volume is referred to as the *reference* and the distorted MRI the *floating* dataset.

The detailed description of the TPS transformation can be found in Bookstein's paper.²⁶ In brief, weighting vector $W = (w_1, w_2, \dots, w_n)$ and the coefficients a_1, a_u, a_v are computed from a series of matrices which are constructed using n pairs of selected control points in the *reference* image (x_i, y_i)

and in the *floating* image (u_i, v_i) . The function transforming a voxel in the *floating* volume to a new coordinate in the *reference* volume is defined as

$$f(u', v') = a_1 + a_u u + a_v v + \sum_{i=0}^n w_i U(|P_i - (u, v)|),$$

where P is the matrix of the coordinates of control points in the reference image and U is a basis function to measure the distance. The computation of the TPS transformation is rather efficient. In our experiment, it took around 5 s to compute a 520×520 -pixel, 8-control point transformation on a Personal Computer (PC) with Intel Pentium® II 400 MHz CPU (Intel Corporation, Sunnyvale, CA) and 256 MB memory.

For comparison purposes, we also implemented the non-deformable registration method reported by Zaider *et al.*¹⁷ and Mizowaki *et al.*²¹ For a particular voxel in the MR space (coordinate z_1), the z coordinate in the US/CT space was obtained from

$$\frac{z_1 - z_{C_1}}{z_{T_1} - z_{B_1}} = \frac{z_2 - z_{C_2}}{z_{T_2} - z_{B_2}},$$

where z_{T_1} and z_{T_2} are the coordinates of the superior aspects of the prostate in the MR and US/CT volume, respectively, z_{B_1} and z_{B_2} refer to the z coordinates of the inferior aspects of the prostate, respectively, and z_{C_1} and z_{C_2} represent the z coordinates of the prostate centroid in the MR space and US/CT space, respectively. Similarly, the (x, y) coordinates were mapped as follows:

$$\frac{y_{A_1} - y_{P_1}}{y_{A_1} - y_1} = \frac{y_{A_2} - y_{P_2}}{y_{A_2} - y_2},$$

$$\frac{x_{L_1} - x_{R_1}}{x_{L_1} - x_1} = \frac{x_{L_2} - x_{R_2}}{x_{L_2} - x_2}.$$

Here, y_A and y_P are the y coordinates of the anterior and posterior aspects of the prostate, respectively, and x_L and x_R refer to, respectively, the x coordinates of the left and right aspects of the prostate. The results obtained by using this method and the newly developed TPS method were quantitatively compared in the phantom and patient studies.

D. Validation of the image registration

For phantom studies, the control points were chosen only in the periphery for the registration of the *floating* and *reference* images. The inserted landmarks were used to trace the displacement and verify the registration accuracy. The control points used in the registration were excluded in computing registration error. After mapping the distorted phantom to the original one, the displacements of the implanted markers were measured with respect to their ideal positions and the mean discrepancy was calculated for each phantom. The mean landmark displacement error (MLDE) was used as a metric for evaluating the quality of the registration.

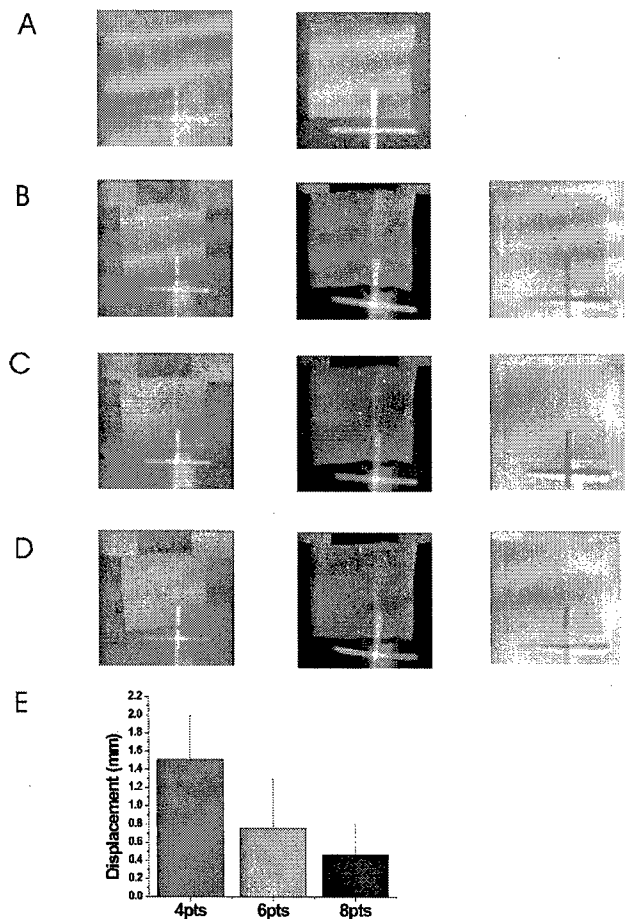


FIG. 2. A registration study using a square phantom deformed by external force. (A) The phantom under the influence of a force (left) and its original shape (right). The distorted phantom is shown in a smaller scale than the original phantom in order to include part of the holder and external object. (B) The position of four control points on the distorted phantom is indicated by pink plus signs (left). The middle shows the computed deformed image. The right column shows the difference image between the computer transformed image and original one. (C) and (D) are similar to B except that six and eight control points are used, respectively. (E) The landmark displacement of the three groups.

For patient studies, typically 6–8 control points were chosen along the contour of the prostate based on the pronounced feature in geometry. Patient MR and CT registration accuracy was estimated by using the centroid position displacement of the prostate and the coincidence index (CI) defined by

$$CI(R, DF) = \frac{DF \cap R}{DF \cup R},$$

where CI is unity when two structures overlap exactly and zero when they are completely disjointed.²⁷ The *deformed floating* (DF) image and the *reference* (R) images were converted to binary for the calculation. The intensity of the voxels inside the contour of the prostate was set 1 and that of other voxels was set 0. The use of CI provided us with an effective measure of the similarity between the warped

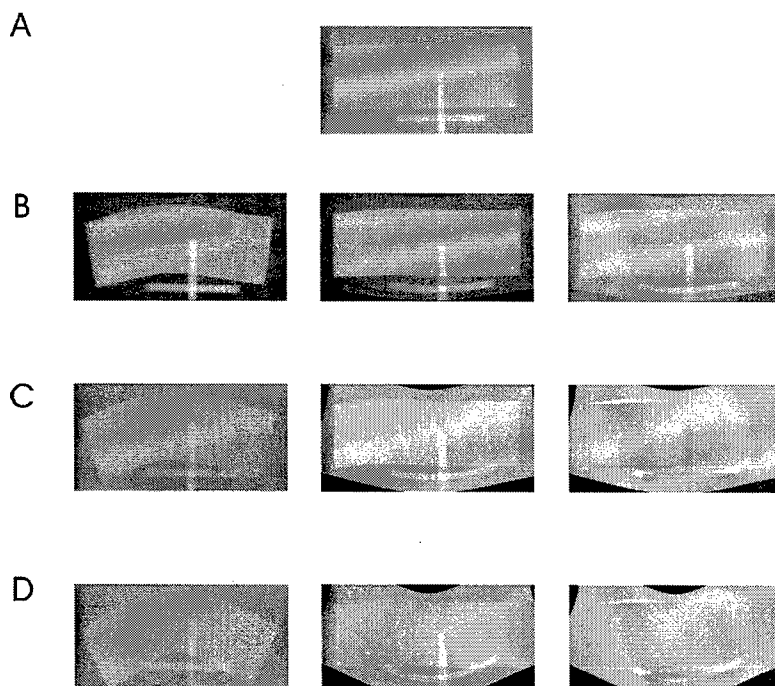


FIG. 3. A registration study by systematically bending a rectangular phantom. (A) The original phantom. (B) The distorted phantom (left), the TPS-warped phantom (middle) and the difference image between computed and original one (right). (C) and (D) are similar to (B) except with increased distortions.

floating (*F*) image and the *reference* (*R*) image. In both evaluations the tissue density was assumed to be homogeneous.

III. RESULTS

A. Phantom studies

We first studied the dependence of registration accuracy on the number of control points. An elastic phantom with dimension $5.5 \times 5.5 \times 1 \text{ cm}^3$ was used here. The phantom was distorted by the insertion of an object in the holder [Fig. 2(A) left] and it restored to the original shape when the object was removed [Fig. 2(A) right]. When four control points were selected along the margin [Fig. 2(B) left], we obtained the warped image shown in the middle column. To evaluate the TPS algorithm, we computed the difference between the TPS predicted and the true image [Fig. 2(B) right panel]. The MLDE was found to be $1.51 \pm 0.49 \text{ mm}$. It is seen that the implanted landmarks do not coincide well in the two images. Next we added two more control points in the periphery and the corresponding mapped image shows reduced registration error with MLDE down to $0.76 \pm 0.54 \text{ mm}$ [Fig. 2(C)]. When eight control points were selected, the MLDE was further reduced to $0.46 \pm 0.34 \text{ mm}$ and no significant landmark displacement was found in the difference image [Fig. 2(D)]. In Fig. 2(E) we summarized the MLDEs when four, six and eight control points were used in the warping calculation. The use of more control points resulted in higher registration accuracy. In practice, however, increasing the number of control points requires additional manual interaction and prolongs the registration process. In the following studies, six to eight pairs of control points were selected for the TPS registration. We also mapped the distorted phantoms onto the *reference* images using a rigid-body registration and scaling

based method.^{17,21} The nondeformable registration resulted in a $2.50 \pm 0.83 \text{ mm}$ MLDE when the maximum displacement was 4.2 mm. Hence, in the situation of a 4.2 mm distortions, the TPS method with eight control points yielded a MLDE that was only 18.4% of the MLDE obtained with the nondeformable model (0.46 vs 2.50 mm).

In the next level of validation we tested the algorithm with a larger rectangular phantom ($9.2 \times 5.1 \times 1 \text{ cm}$) to allow more flexible distortions. The *reference* image is shown in Fig. 3(A). The *floating* images under a few different levels of distortion are shown in the left columns of B, C, and D. Eight control points were used here to register the *floating* and *reference* images. The middle panels of B, C, and D show the results after the TPS transformation. The differences between the TPS predictions and *references* are shown on the right panels of B, C, and D. The quality of the TPS mapping was assessed by using the maximum and mean landmark displacement error. As summarized in Table I, for the distortion shown in Fig. 3(A), the nondeformable registration gave a MLDE of $4.62 \pm 2.71 \text{ mm}$, whereas the deformable registration reduced the error down to $0.45 \pm 0.53 \text{ mm}$. For the studies shown in Figs. 3(C) and 3(D), the nondeformable registration yielded MLDEs to $7.35 \pm 4.20 \text{ mm}$ and $12.95 \pm 6.57 \text{ mm}$, respectively. The maximum displacement errors in some fiducials are larger than 10 mm. The application of the deformable warping module significantly improved the mapping and led to MLDEs of $0.57 \pm 0.49 \text{ mm}$ and $0.62 \pm 0.39 \text{ mm}$, respectively. The total displacements of control points to bring about the registration are 38.43, 58.76, and 102.34 mm, respectively, in these cases (Table I). The registration errors increase with them. The largest registration error between the TPS prediction and the ideal situation were found to be 1.09 mm, 1.05 mm, and 0.99 mm, respectively, for the three phantom distortions.

TABLE I. Deformable and nondeformable registration error of the distorted phantoms shown in Figs. 3(B)–3(D). Total displacements of control points (CPs) to bring about the registration are listed. The maximum refers to the maximum landmark displacement error and the mean refers to the mean landmark displacement error in each case.

Distortion (mm)	Total displacements of CPs	Nondeformable registration		Deformable registration	
		Maximum	Mean	Maximum	Mean
Fig. 3(B)	38.43	9.83	4.62±2.71	1.09	0.45±0.53
Fig. 3(C)	58.76	14.74	7.35±4.20	1.05	0.57±0.49
Fig. 3(D)	102.34	23.07	12.95±6.57	0.99	0.62±0.39

Considering the severity of the distortion in this example and the voxel size of MRSI is of 5 mm in dimension, the TPS based method seems adequate to correlate the functional data onto the corresponding voxels in CT.

In order to examine the consistency of the registration, we inverted the previous transformation procedure by transforming the TPS-warped images [Figs. 3(B)–3(D) middle panels] back to the distorted *floating* images. The calculation results are shown in the left columns of Figs. 4(A)–4(C). The original deformed phantom images (the left columns of Fig. 3)

are subtracted from them and are shown in the right panels. The resultant overlap of fiducial points was excellent in all three cases, suggesting the TPS is capable of generating consistent good mapping independent of the starting images. The MLDEs for the three groups were 0.23 ± 0.08 , 0.23 ± 0.18 , and 0.20 ± 0.11 mm, respectively. Maximum landmark registration discrepancies were found to be 0.35, 0.50, and 0.31 mm, respectively [Fig. 4(D)].

B. Patient studies

We first studied where the distortion most likely happens in the ER-based MR images. After target segmentation and rotation operation, we compared the dimensions of the prostate in the datasets of three patients. The height of the prostate along superior-posterior axis was found almost the same (3.1% discrepancy) in CT and ER based-MR images. The width along left-right axis and the length along anterior-posterior axis differ a lot between CT and MR images. We measured the width and length of the prostate in the middle axial slices of three patients. The width of the prostate in MRI is $(115.0 \pm 1.7)\%$ of that of the prostate in CT. The length of the prostate in MRI is $(83.3 \pm 1.5)\%$ of that of the prostate in CT. After further comparing the shape of the prostates in CT and ER-MRI, we conclude that distortion mostly happens in the transverse plane. This observation suggests that, to a reasonable approximation, we could perform the mapping procedure in a slice by slice fashion.

The TPS transformation was applied to the coregistration of the CT and ER-based MR images. We show a representative axial slice of a patient's CT images in Fig. 5(A). To have a better view of the volume of interest, we selected a rectangular region encompassing the prostate [Fig. 5(A) right]. The MR images were acquired with high resolution and the posterior portion of the image was distorted by the presence of the ER coil [Fig. 5(B) left]. Eight control points were chosen along the contour of the corresponding MR and CT images. TPS transformation was applied to the distorted MR image and the mapped MRI contour of the prostate overlapped almost completely with that from the CT scan [Fig. 5(B) right]. Difference between TPS-derived MR contour and CT contour is shown in Fig. 5(C). Most prostate regions were in good agreement including the seriously contorted left and right posterior regions of the image. Similar results were obtained from the other patient.

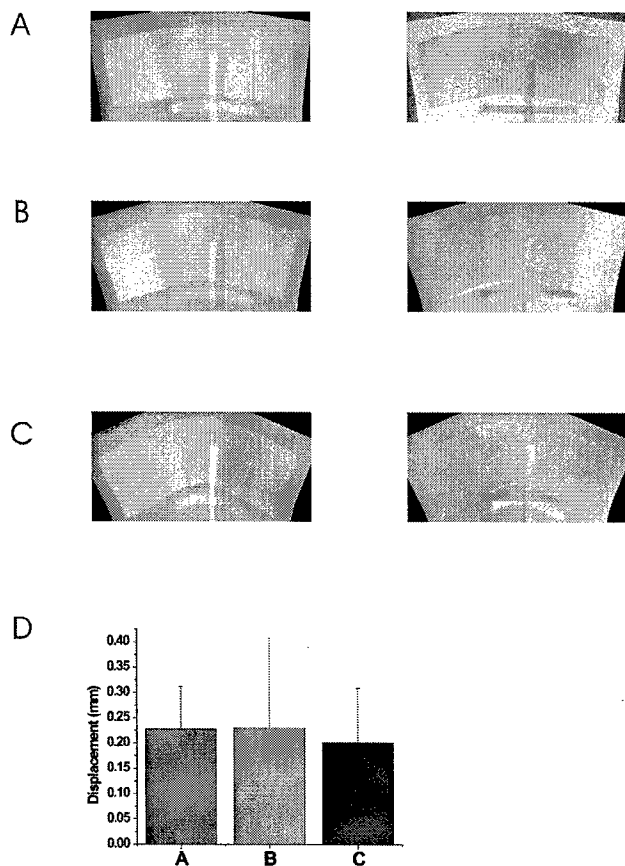


FIG. 4. Registration consistency test. Left panels of (A), (B), and (C) represent the computer-warped images with the middle panel images of Figs. 3(B)–3(D) as input. Right panels of (A), (B), and (C) represent the corresponding difference images between the mapped and the original images. (D), Landmark displacement between the model prediction and the actual position for the three groups.

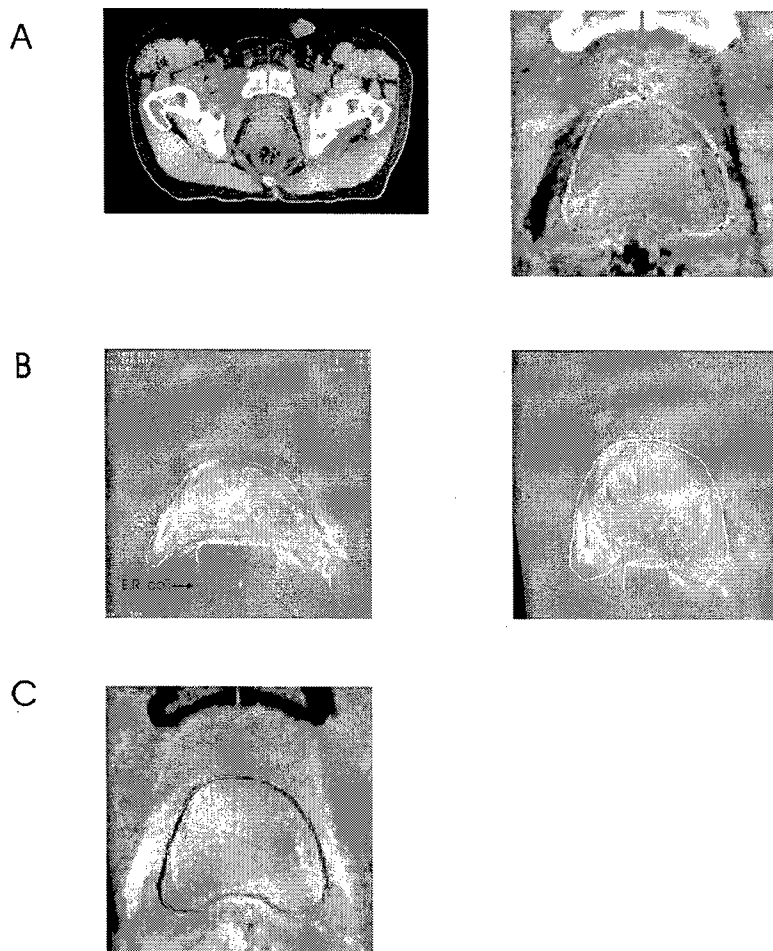


FIG. 5. Deformable registration of the prostate gland in a patient. (A) A transverse CT study (left) and a region of interest encompassing the prostate (right). The control points are denoted with plus signs. (B) The MRI study (left) and transformed image (right) by TPS based registration. The control points are denoted with plus signs. (C) Difference between the CT and the mapped MR image.

We used centroid position displacement and coincidence index (CI) between the mapped MR images and the CT images to quantify the registration accuracy for three datasets. Using the TPS method, the centroid displacement was 0.56 ± 0.09 mm, significantly less than that of the nondeformable registration [2.03 ± 0.38 mm, Fig. 6(A)]. The CI indices were found to be close to unity ($93.1 \pm 5.0\%$), indicating that the TPS algorithm is able to model the nonrigid soft tissue deformation caused by the endorectal coil placement [Fig. 6(B)]. On the other hand, a much lower CI, ($49.5 \pm 8.9\%$), was found when using the nondeformable registration. This suggests that fusion with a rigid-body transformation and scaling is inadequate to deal with the system involving the images acquired with the ER coils.

The registration error depends on the appropriate placement of control points. We studied registration inconsistencies between different trials of one operator and between three operators. The intraoperator experiment was repeated five times on one patient's data. The centroid displacement was found to be in the range of 0.31–0.65 mm. The CI indices were found to be from 91.7% to 93.5%. Three operators were asked to repeat the control point placement five times and the mean results between the operators were compared. The centroid displacements were found to be 0.55 ± 0.30 , 0.47 ± 0.17 , and 0.54 ± 0.25 mm, respectively.

The CI indices, corresponding to three operators, were ($92.7 \pm 0.9\%$), ($93.5 \pm 0.7\%$), and ($92.5 \pm 1.2\%$), respectively. The centroid displacement and CI index show no significant difference between trials and operators.

IV. DISCUSSION AND CONCLUSION

Registration has been implemented in several commercial medical image analysis and radiation treatment planning systems. For example, Radionics (Radionics™, Burlington, MA) has developed ImageFusion software which provides the ability to fuse multiple image sets based on the mutual information. AcQSim Oncodiagnostic Simulation/Localization System (Philips Medical Systems, Cleveland, OH) provides two registration methods: point matching (a minimum of three common points need to be selected on both sets of images registered) and interactive image-based registration (a color wash of one image set is displayed over a gray scale image of the other). At this point, they all use a rigid-body transformation and scaling, which maintain the straightness of lines, and hence cannot accommodate contour/shape distortion. In reality, the shape of the prostate gland can be easily changed by many factors such as patient position change, invasive brachytherapy procedures or endorectal coil placement during high resolution MR/MRS im-

ages acquisition.²⁸ To help physicians to segment the prostate gland and possible intraprostatic lesions by incorporating MRI/MRSI metabolic data on a CT-based treatment planning system, there is an indisputable need for developing a computationally efficient deformable registration technique to achieve voxel to voxel mapping. In this work, we used a TPS method to register the endorectal coil-based MR data with CT images. The data presented in the last section suggests that the TPS technique is well suited for this type of application.

The warping process was carried out in a 2D slice-by-slice and is worthy of further investigation. This may result in the registration error in the longitudinal direction. Based on our observations for the three patients involved in this study, it seems that the distortion occurs mainly along the right-left and the anterior-posterior directions. The height of the prostate along the superior-inferior axis changes insignificantly in the MR and CT datasets. This is consistent with the finding by another group in 1.5 T MR imaging of the prostate.²⁸ In actuality, it is possible to extend the current quasi-3D model to a fully 3D one. The current study sheds useful insight into this type of extension and provides a natural starting point for the implementation of a complete 3D TPS mapping. We are aware that the prostate volume drawn from CT may be different from that in MR.²⁹ Currently we simply rely on the expert opinions from the radiation oncologist. Hopefully, with the common efforts from physicists and physicians, this difficult issue will be resolved in the near future.

A few more sophisticated deformable registration methods have been investigated by several groups. A viscous-fluid transformation and fluid-landmark registration technique have been proposed to model the nonrigid deformation of organs in intracavitary brachytherapy.^{27,30} A finite-element method has been used to model the tissue mechanical property and to register brain and prostate images.^{31,32} A biomechanical model of an elastic body has been used to quantify patient organ motion in the process of radiation therapy so that the dose delivered on the volume of a deforming organ can be accumulated.³³ These methods are usually computationally intensive. Moreover, the model parameters normally need to be determined empirically because of the lack of tissue biomechanical data in the literature, which compromises the advantages of these physics-based models. More recently, some registration schemes are designed to match both control point and intensity information.³⁴⁻³⁶ In addition to guaranteeing a one to one correspondence of the prescribed control points, they minimize an intensity based measure for the remaining parts of the images. The method should yield higher accuracy than the control point only based method and should be more computationally efficient than the intensity only based method. In addition to computation efficiency, there are another two considerations for us to use the TPS based registration. Firstly, our ultimate goal is to map the metabolic data of MRSI to the correct voxels in CT. The voxel size of MRSI is generally very coarse (currently ~5 mm) compared to the anatomic images. The increased registration accuracy by more advanced methods

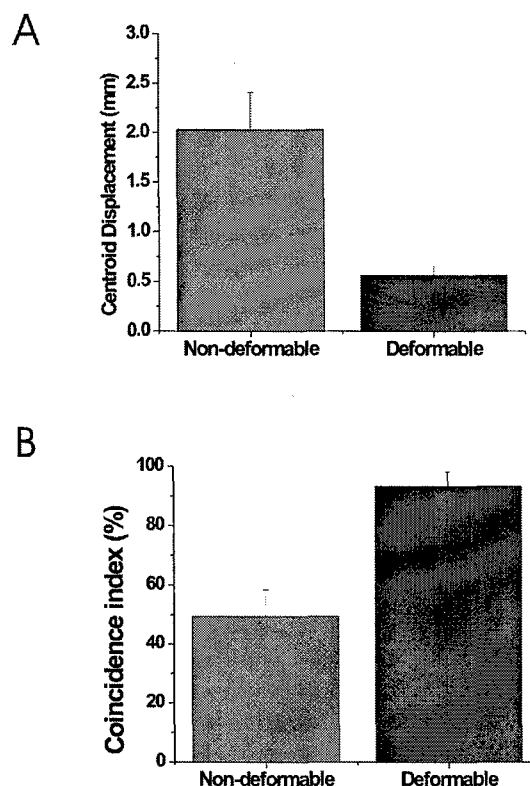


FIG. 6. Centroid position displacement (A) and coincidence index of deformable and nondeformable registration in patient studies (B).

may be not necessary for this mapping purpose. Secondly, our registration target is the prostate only which is of similar intensity. The linear interpolation after matching control points in TPS should be sufficiently accurate for registering a uniform small structure. The phantom and patient studies have shown that the TPS approach is computationally efficient and can yield clinically acceptable registration accuracy for our purpose.

It is noted that the TPS based registration needs manual placement of control points, which requires the input from an experienced clinician. This is similar to the previously reported rigid body-based registration method.^{17,21} The intra- and interoperator experiments have shown that the variances of results are small and registration accuracy does not depend on the different operators or different trials significantly.

In conclusion, we have implemented a TPS transformation algorithm to map voxels in endorectal coil-based prostate MR/MRS images with those in CT images. The deformable mapping technique significantly improved the previously reported nondeformable method and should be adequate for routine clinical application. The accuracy of the approach has been tested by using phantom and patient studies. The registration scheme should be useful to map the functional MRSI data onto CT to guide the design of conformal radiation treatment plans.

ACKNOWLEDGMENTS

This work was presented in the 45th ASTRO annual meeting in 2003. We would like to thank the assistance of A.

Boyer, C. Cardenas, F. van den Haak, and Y. Yang from Stanford University. We also appreciate the enlightening discussions with S. Chang and L. Goyal from The University of North Carolina at Chapel Hill, and W. Du from The University of Chicago. This work was in part supported by a research grant from the prostate cancer research program of U.S. Department of Defense (DAMD17-03-1-0023).

^aPresent address: Department of Radiation Oncology, The University of North Carolina, 101 Manning Dr., Chapel Hill, North Carolina 27599-7512. E-mail: Jun_Lian@med.unc.edu; Phone: (919) 966 1101; Fax: (919) 966 7681.

^bAuthor to whom correspondence should be addressed. E-mail: lei@reyes.stanford.edu; Phone: (650) 498 7896; Fax: (650) 498 4015.

¹C. Bartolozzi, L. Crocetti, I. Menchi, S. Ortori, and R. Lencioni, "Endorectal magnetic resonance imaging in local staging of prostate carcinoma," *Abdom. Imaging* **26**, 111–122 (2001).

²M.D. Schnall, Y. Imai, J. Tomaszewski, H.M. Pollack, R.E. Lenkinski, and H.Y. Kressel, "Prostate cancer: Local staging with endorectal surface coil MR imaging," *Radiology* **178**, 797–802 (1991).

³M.D. Schnall, R.E. Lenkinski, H.M. Pollack, Y. Imai, and H.Y. Kressel, "Prostate: MR imaging with an endorectal surface coil," *Radiology* **172**, 570–574 (1989).

⁴J.F. Martin, P. Hajek, L. Baker, V. Gyls-Morin, R. Fitzmorris-Glass, and R.R. Mattrey, "Inflatable surface coil for MR imaging of the prostate," *Radiology* **167**, 268–270 (1988).

⁵R.A. Huch Boni, C. Meyenberger, J. Pok Lundquist, F. Trinkler, U. Lutolf, and G.P. Krestin, "Value of endorectal coil versus body coil MRI for diagnosis of recurrent pelvic malignancies," *Abdom. Imaging* **21**, 345–352 (1996).

⁶Y. Kaji, J. Kurhanewicz, H. Hricak, D.L. Sokolov, L.R. Huang, S.J. Nelson, and D.B. Vigneron, "Localizing prostate cancer in the presence of postbiopsy changes on MR images: Role of proton MR spectroscopic imaging," *Radiology* **206**, 785–790 (1998).

⁷J. Kurhanewicz, D.B. Vigneron, H. Hricak, P. Narayan, P. Carroll, and S.J. Nelson, "Three-dimensional H-1 MR spectroscopic imaging of the *in situ* human prostate with high (0.24–0.7-cm³) spatial resolution," *Radiology* **198**, 795–805 (1996).

⁸L. Kwok, J.K. Smith, M. Castillo, M.G. Ewend, S. Cush, T. Hensing, M. Varia, D. Morris, and T.W. Bouldin, "Clinical applications of proton MR spectroscopy in oncology," *Technol. Cancer Res. Treatment* **1**, 17–28 (2002).

⁹J.C. Presti, Jr., H. Hricak, P.A. Narayan, K. Shinohara, S. White, and P.R. Carroll, "Local staging of prostatic carcinoma: comparison of transrectal sonography and endorectal MR imaging," *AJR, Am. J. Roentgenol.* **166**, 103–108 (1996).

¹⁰H. Hricak, G.C. Doms, R.B. Jeffrey, A. Avallone, D. Jacobs, W.K. Benton, P. Narayan, and E.A. Tanagho, "Prostatic carcinoma: Staging by clinical assessment, CT, and MR imaging," *Radiology* **162**, 331–336 (1987).

¹¹M. Perrotti, R.P. Kaufman, Jr., T.A. Jennings, H.T. Thaler, S.M. Soloway, M.D. Rifkin, and H.A. Fisher, "Endo-rectal coil magnetic resonance imaging in clinically localized prostate cancer: Is it accurate?," *J. Urol. (Baltimore)* **156**, 106–109 (1996).

¹²G.J. Jager, J.L. Severens, J.R. Thornbury, J.J. de La Rosette, S.H. Ruijs, and J.O. Barentsz, "Prostate cancer staging: Should MR imaging be used?—A decision analytic approach," *Radiology* **215**, 445–451 (2000).

¹³S.F. Quinn, D.A. Franzini, T.A. Demlow, D.R. Rosencrantz, J. Kim, R.M. Hanna, and J. Szumowski, "MR imaging of prostate cancer with an endorectal surface coil technique: correlation with whole-mount specimens," *Radiology* **190**, 323–327 (1994).

¹⁴P. Xia, B. Pickett, E. Vigneault, L.J. Verhey, and M. Roach, 3rd, "Forward or inversely planned segmental multileaf collimator IMRT and sequential tomotherapy to treat multiple dominant intraprostatic lesions of prostate cancer to 90 Gy," *Int. J. Radiat. Oncol., Biol., Phys.* **51**, 244–254 (2001).

¹⁵S.J. DiBiase, K. Hosseinzadeh, R.P. Gullapalli, S.C. Jacobs, M.J. Naslund, G.N. Sklar, R.B. Alexander, and C. Yu, "Magnetic resonance spectroscopic imaging-guided brachytherapy for localized prostate cancer," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 429–438 (2002).

¹⁶L. Xing, C. Cotrutz, S. Hunjan, A.L. Boyer, E. Adalsteinsson, and D. Spielman, "Inverse planning for functional image-guided intensity-

modulated radiation therapy," *Phys. Med. Biol.* **47**, 3567–3578 (2002).

¹⁷M. Zaider, M.J. Zelefsky, E.K. Lee, K.L. Zakian, H.I. Amols, J. Dyke, G. Cohen, Y. Hu, A.K. Endi, C. Chui, and J.A. Koutcher, "Treatment planning for prostate implants using magnetic-resonance spectroscopy imaging," *Int. J. Radiat. Oncol., Biol., Phys.* **47**, 1085–1096 (2000).

¹⁸C.C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J.A. Koutcher, "Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality," *Int. J. Radiat. Oncol., Biol., Phys.* **47**, 551–560 (2000).

¹⁹J. Lian, S. Hunjan, C. Dumoulin, J. Levin, R. Watkins, K. Rohling, R. Giaquinto, D. Kim, A. Lo, D. Spielman, B. Daniel, and L. Xing, "Integrating deformable MRI/MRSI and CT image registration into the prostate IMRT treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **57**, S207 (2003).

²⁰X. Wu, C.X. Yu, S.J. DiBiase, and R. Gullapalli, "The application of deformable image registration for MRS in prostate treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **57**, S207–S208 (2003).

²¹T. Mizowaki, G.N. Cohen, A.Y. Fung, and M. Zaider, "Towards integrating functional imaging in the treatment of prostate cancer with radiation: The registration of the MR spectroscopy imaging to ultrasound/CT images and its implementation in treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **54**, 1558–1564 (2002).

²²B. Fei, A. Wheaton, Z. Lee, J.L. Duerk, and D.L. Wilson, "Automatic MR volume registration and its evaluation for the pelvis and prostate," *Phys. Med. Biol.* **47**, 823–838 (2002).

²³B. Fei, C. Kemper, and D.L. Wilson, "A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes," *Comput. Med. Imaging Graph.* **27**, 267–281 (2003).

²⁴K.M. Brock, J.M. Balter, L.A. Dawson, M.L. Kessler, and C.R. Meyer, "Automated generation of a four-dimensional model of the liver using warping and mutual information," *Med. Phys.* **30**, 1128–1133 (2003).

²⁵S. Fantozzi, A. Cappello, and A. Leardini, "A global method based on thin-plate splines for correction of geometric distortion: An application to fluoroscopic images," *Med. Phys.* **30**, 124–131 (2003).

²⁶F.L. Bookstein, "Principal Warps: Thin Plate Splines and the Decomposition of Deformations," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 567–585 (1989).

²⁷G.E. Christensen, B. Carlson, K.S. Chao, P. Yin, P.W. Grigsby, K. Nguyen, J.F. Dempsey, F.A. Lerma, K.T. Bae, M.W. Vannier, and J.F. Williamson, "Image-based dose planning of intracavitary brachytherapy: Registration of serial-imaging studies using deformable anatomic templates," *Int. J. Radiat. Oncol., Biol., Phys.* **51**, 227–243 (2001).

²⁸M. Hirose, A. Bharatha, N. Hata, K.H. Zou, S.K. Warfield, R.A. Cormack, A. D'Amico, R. Kikinis, F.A. Jolesz, and C.M. Tempany, "Quantitative MR imaging assessment of prostate gland deformation before and during MR imaging-guided brachytherapy," *Acad. Radiol.* **9**, 906–912 (2002).

²⁹C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J.V. Lebesque, "Definition of the prostate in CT and MRI: A multiobserver study," *Int. J. Radiat. Oncol., Biol., Phys.* **43**, 57–66 (1999).

³⁰S.C. Joshi and M.I. Miller, "Landmark matching via large deformation diffeomorphisms," *IEEE Trans. Image Process.* **9**, 1357–1370 (2000).

³¹A. Bharatha, M. Hirose, N. Hata, S.K. Warfield, M. Ferrant, K.H. Zou, E. Suarez-Santana, J. Ruiz-Alzola, A. D'Amico, R.A. Cormack, R. Kikinis, F.A. Jolesz, and C.M. Tempany, "Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging," *Med. Phys.* **28**, 2551–2560 (2001).

³²M. Ferrant, A. Nabavi, B. Macq, F.A. Jolesz, R. Kikinis, and S.K. Warfield, "Registration of 3-D intraoperative MR images of the brain using a finite-element biomechanical model," *IEEE Trans. Med. Imaging* **20**, 1384–1397 (2001).

³³D. Yan, D.A. Jaffray, and J.W. Wong, "A model to accumulate fractionated dose in a deforming organ," *Int. J. Radiat. Oncol., Biol., Phys.* **44**, 665–675 (1999).

³⁴B. Fischer and J. Modersitzki, *Intensity Based Image Registration With a Guaranteed One-to-One Point Match*, Advances in Biomedical Image Analysis, Selected Papers from the German BVM-Workshop on Medical Image Processing, Special Issue, Methods of Information in Medicine, edited by T. Tolxdorff (Schattauer Verlag, Stuttgart, 2003).

³⁵A. Fischer and J. Modersitzki, "Combining landmark and intensity driven registrations," *PAMM* **3**, 32–35 (2003).

³⁶J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Trans. Image Process.* **12**, 1427–1442 (2003).

Towards biologically conformal radiation therapy (BCRT): Selective IMRT dose escalation under the guidance of spatial biology distribution

Yong Yang and Lei Xing^{a)}

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305-5847

(Received 28 September 2004; revised 17 March 2005; published 11 May 2005)

It is well known that the spatial biology distribution (e.g., clonogen density, radiosensitivity, tumor proliferation rate, functional importance) in most tumors and sensitive structures is heterogeneous. Recent progress in biological imaging is making the mapping of this distribution increasingly possible. The purpose of this work is to establish a theoretical framework to quantitatively incorporate the spatial biology data into intensity modulated radiation therapy (IMRT) inverse planning. In order to implement this, we first derive a general formula for determining the desired dose to each tumor voxel for a known biology distribution of the tumor based on a linear-quadratic model. The desired target dose distribution is then used as the prescription for inverse planning. An objective function with the voxel-dependent prescription is constructed with incorporation of the nonuniform dose prescription. The functional unit density distribution in a sensitive structure is also considered phenomenologically when constructing the objective function. Two cases with different hypothetical biology distributions are used to illustrate the new inverse planning formalism. For comparison, treatments with a few uniform dose prescriptions and a simultaneous integrated boost are also planned. The biological indices, tumor control probability (TCP) and normal tissue complication probability (NTCP), are calculated for both types of plans and the superiority of the proposed technique over the conventional dose escalation scheme is demonstrated. Our calculations revealed that it is technically feasible to produce deliberately nonuniform dose distributions with consideration of biological information. Compared with the conventional dose escalation schemes, the new technique is capable of generating biologically conformal IMRT plans that significantly improve the TCP while reducing or keeping the NTCPs at their current levels. Biologically conformal radiation therapy (BCRT) incorporates patient-specific biological information and provides an outstanding opportunity for us to truly individualize radiation treatment. The proposed formalism lays a technical foundation for BCRT and allows us to maximally exploit the technical capacity of IMRT to more intelligently escalate the radiation dose. © 2005 American Association of Physicists in Medicine. [DOI: 10.1118/1.1924312]

Key words: inverse planning, biological model, TCP, NTCP, IMRT

I. INTRODUCTION

Intensity modulated radiation therapy (IMRT) has been used clinically to provide a highly conformal radiation dose to the target volume while reducing the doses to the surrounding sensitive structures.¹⁻¹³ The current IMRT inverse planning is typically aimed at producing a homogeneous target dose under the assumption of uniform biology within the target volume. In reality, it is known that the spatial biology distributions in most tumors and normal tissues are rarely homogeneous. To maximize the efficacy of IMRT, it is desirable to take the inhomogeneous biological information into account and to produce customized nonuniform dose distributions on a patient specific basis. This type of radiation treatment is referred to as biologically conformal radiotherapy (BCRT).¹⁴⁻¹⁹ The simultaneous integrated boost (SIB) to elective volumes recently appearing in the literature^{11,17,20} represents a simple example of BCRT. However, an underlying deficiency of the current SIB approach is that the boost doses are based on previous experience, not patient-specific biological information characterizing the spatial tumor burden distribution.

To establish the BCRT treatment planning scheme, three major aspects must be addressed: (i) Determination of the distribution of biological properties of the tumor and critical structures; (ii) Prescription of the desired dose distribution for inverse planning; and (iii) Inverse planning to generate most faithfully the prescribed nonuniform dose distribution. Recently spurred efforts in biological imaging, such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and magnetic resonance spectroscopy imaging (MRSI), are aimed at providing solutions to the first problem.²¹⁻³¹ To give a few examples, the clonogen density in malignant glioma can be obtained based on the choline/creatine ratio through MRSI,^{29,30} tumor hypoxia can be quantified using PET imaging with fluorinated misonidazole (FMISO),^{27,28} tumor proliferation rate can be obtained based on the voxel activity level in DNA proliferation imaging (e.g., fluoro-L-thymidine PET),^{25,26,32} and lung functional importance distributions can be obtained by perfusion imaging.³³ While the development of molecular imaging techniques is critically important in mapping out biology distributions, the successful integration of this information

into IMRT planning through steps (ii) and (iii) is also indispensable to fully exploit the obtained biology information to improve patient care. In this study we focus our efforts on the last two problems, with the optimistic assumption that spatial biology distributions within a patient have already been determined from biological imaging or other means. Our goal is to establish a theoretical framework for quantitatively incorporating the biological data into IMRT inverse optimization, and to show the advantage of the selective dose escalation scheme in enhancing tumor control probability (TCP) and reducing the normal tissue complication probability (NTCP). In conjunction with the rapid development of molecular imaging techniques, this study lays a technical foundation for BCRT and provides a basis for clinically realizing the new treatment strategy in the future.

II. METHODS AND MATERIALS

A. Biological characterization and nonuniform target dose prescription

We assume that biological properties influencing radiation treatment are characterized phenomenologically by three radiobiology parameters: clonogen density (ρ), radiosensitivity (α), and proliferation rate (γ). Generally, these parameters are voxel dependent. In this work we concentrate on their spatial variation within tumor, and ignore the time dependence of the last two parameters.

To accomplish BCRT, an important step is to derive the desired dose distribution that maximizes the cell killing based on (ρ, α, γ) metrics. In the case of uniform biology, it is well known that the target dose should be uniformly distributed. It is, however, not clear at all what form of dose distribution should be used to maximize the cell killing for an arbitrary biology distribution. We start from a linear quadratic (LQ) model³⁴⁻³⁶ with inclusion of the tumor cell proliferation. According to this model, the tumor clonogen survival S_i in a voxel of volume V_i after an irradiating dose D_i is given by

$$S_i = \rho_i V_i \exp(-\alpha'_i D_i + \gamma_i \Delta T), \quad (1)$$

where $\alpha'_i = \alpha_i [1 + d_i / (\alpha_i / \beta_i)]$, ρ_i is the initial clonogen density, d_i is the fractional dose, α_i and β_i are the linear-quadratic coefficients of the cell survival curve, $\gamma_i = \ln 2 / T_p$ is the cell proliferation rate, T_p is the potential cell doubling time, and ΔT is the overall treatment time. The TCP of a voxel i can be expressed as

$$\text{TCP}_i = \exp[-\rho_i V_i \exp(-\alpha'_i D_i + \gamma_i \Delta T)]. \quad (2)$$

The TCP for the whole tumor is the product of the TCP_i of all voxels within the tumor volume, i.e.,

$$\text{TCP} = \prod_i \text{TCP}_i. \quad (3)$$

For a given set of $\{\rho, \alpha, \gamma\}$, the task is to find the dose distribution that maximizes the TCP. Because of the limitation of normal tissue dose tolerances, an arbitrarily high dose to the tumor cannot be achieved and certain constraints need to be imposed.³⁶⁻⁴¹ In line with previous researchers,^{36,37,40} we

restrict the integral dose to the tumor volume to a constant. Mathematically, the constraint is written as

$$\sum_i m_i D_i = E_t, \quad (4)$$

where m_i is the mass of voxel i , and E_t is the integral target dose.

With the above formulation, the task becomes the maximization of the TCP under the constraint (4). The Lagrange multiplier method is employed to solve the problem. In this approach, a function

$$L(\text{TCP}_1, \dots, \text{TCP}_i, \dots) = \prod_i \text{TCP}_i + \lambda \left(\sum_i m_i D_i - E_t \right), \quad (5)$$

is introduced, where λ is the Lagrange multiplier, and the solution is obtained by solving the equations

$$\frac{\partial L}{\partial \text{TCP}_i} = 0. \quad (6)$$

When mass and volume are equal for all voxels in the target, using a process similar to Ebert and Hoban⁴⁰ (see the Appendix), we obtained a general formula for determining the desired dose, $D_0^T(i)$, at the voxel i

$$D_0^T(i) = \frac{\alpha'_{\text{ref}}}{\alpha'_i} D_{\text{ref}} - \frac{1}{\alpha'_i} (\gamma_{\text{ref}} - \gamma_i) \Delta T - \frac{1}{\alpha'_i} \ln \left(\frac{\alpha'_{\text{ref}} \rho_{\text{ref}}}{\alpha'_i \rho_i} \right), \quad (7)$$

where D_{ref} is the reference dose for the voxel with reference radiobiological parameters ($\rho_{\text{ref}}, \alpha_{\text{ref}}, \gamma_{\text{ref}}$). In general, D_{ref} should be set to a value that yields a clinically sensible TCP at the reference voxel. For a given disease site, the radiation dose used in current clinical practice with "intent to cure" can be used as a good starting point in selecting the value of D_{ref} . Using Eq. (7), it is straightforward to determine the desired target prescription dose once the radiobiological parameter (ρ, α, γ) metrics and D_{ref} are known. Note that the desired dose distribution represents an ideal situation without considering the specific dosimetric tolerances of the sensitive structures. In reality, this dose distribution may or may not be exactly realizable. Nevertheless, it sets a landmark and serves as the prescription dose in inverse planning to guide the dose optimization process.

The fractional dose, d_i , is required to obtain the parameter α'_i in Eq. (7). On the other hand, d_i is not known until $D_0^T(i)$ is known. We use a simple iterative method to solve the dilemma. First, the fractional dose is initially set to $d_i = D_{\text{ref}} / N_f$, N_f being the fractional number. Second, $D_0^T(i)$ is calculated using Eq. (7) and $d_i = D_0^T(i) / N_f$ is updated. The new $D_0^T(i)$ is then obtained using the updated d_i . We find that $D_0^T(i)$ converges to the solution in less than five iterations. In this study we set $\alpha / \beta = 10$ Gy for all target voxels. The formalism proposed here is, however, general and can be extended to deal with nonuniform distributions of the α / β ratio.

B. Inverse planning with spatially nonuniform dose prescription

The next logical step after obtaining the calculated prescription dose is to use inverse planning to derive the optimal beam profiles that will produce the prescribed dose distribution. To proceed, we construct an objective function to take the known biological information into account. In addition to the voxel-specific prescription as determined by Eq. (7), the nonlinear dose responses of tumor and normal structures are considered using the concept of equivalent volume⁴²⁻⁴⁸ of a voxel, which is defined as

$$(\Delta V_{\text{eff}})_i = V_i \phi(i) (D(i)/D_i)^{1/n}, \quad (8)$$

where $(\Delta V_{\text{eff}})_i$ is the effective volume for voxel i with volume V_i and dose $D(i)$, D_i is the desired dose for a target voxel or the $\text{TD}_{5/5}$ of the corresponding organ, and $\phi(i)$ is the functional unit density. The value of n characterizes the dose-volume effect of an organ and reflects its architecture (serial or parallel) of the sensitive structure. It is obtained by fitting to clinical dose-volume data. For a sensitive structure, n is a positive number ($n > 0$) while for a target, n should be assigned with a small negative value ($-1 < n < 0$). $\phi(i) \equiv 1$ for a target voxel.

A general form of the inverse planning objective function in the voxel domain is written as

$$F = \sum_{\tau=1}^{t_{\tau}} r_{\tau} \frac{1}{N_{\tau}} \sum_{i=1}^{N_{\tau}} \{1 + [D_c(i)/D_0^{\tau}(i)]^{1/n_{\tau}}\} [D_c(i) - D_0^{\tau}(i)]^2 \\ + \sum_{\sigma=1}^{s_{\sigma}} r_{\sigma} \frac{1}{N_{\sigma}} \sum_{i=1}^{N_{\sigma}} \{1 + \phi_{\sigma}(i) [D_c(i)/\text{TD}_{\sigma,5/5}]^{1/n_{\sigma}}\} D_c(i)^2, \quad (9)$$

where r_{τ} and r_{σ} are the structure specific importance factors of target τ and sensitive structure σ , respectively, t_{τ} and s_{σ} the number of targets and sensitive structures, N_{τ} and N_{σ} the total number of voxels of target τ or sensitive structure σ , n_{τ} and n_{σ} the n parameter of target τ and sensitive structure σ , $D_c(i)$ the calculated dose in voxel i , $D_0^{\tau}(i)$ the prescription dose in a target voxel i given by Eq. (7), and $\text{TD}_{\sigma,5/5}$ the $\text{TD}_{5/5}$ of sensitive structure σ . The objective function becomes the conventional quadratic objective function if the term in the bracket inside each summation is set to unity (this is true when the dose-volume effect is negligible, i.e., when $n_{\tau}=n_{\sigma}=+\infty$). More detailed information about the optimization algorithm can be found in Ref. 49.

C. Implementation

A software module for optimizing the objective function (9) is implemented in the platform of the PLUNC treatment planning system (University of North Carolina, Chapel Hill, NC). The dose calculation engine and a variety of image/beam/plan display and evaluation tools of PLUNC are used to review and compare the optimization results. The ray-by-ray iterative algorithm (SIITP) reported earlier^{50,51} is employed to obtain the optimal beam intensity profiles. The dose-volume histograms (DVHs) of the involved organs are dis-

TABLE I. Dose-volume parameters of various sensitive structures used for calculating NTCP in this study.

Sensitive structures	n	m	$D_{50/5}$ (Gy)
Bladder	0.50	0.11	80
Rectum	0.12	0.15	80
Femoral head	0.25	0.12	65

played at the end of each iterative step to visually monitor the optimization process.

D. Plan review tools

It is desirable to extend the currently used plan review tools to deal with a biologically heterogeneous system. For a target, we define the effective dose at a voxel as the physical dose normalized by the desired dose determined by Eq. (7). The effective-dose volume histogram (EDVH), which is obtained by replacing the dose with the effective dose in conventional DVH, is a useful tool for assessing BCRT plans. For a sensitive structure we replace the fractional volume by $\phi_i V_i$ to construct a functional dose volume histogram (FDVH), similar to that proposed by Lu *et al.*⁵² and Marks *et al.*³³ After including the heterogeneous biological information into the EDVH or FDVH, the wisdom used in interpreting a conventional DVH can be applied to assess the BCRT plans. In addition to the effective dose and the EDVH or FDVH, a cluster of DVHs, each corresponding to a given set of biological parameters $\{\rho, \alpha, \gamma\}$, is also useful to assess dosimetric behavior of the system as a function of the biological status of the system.

Besides the dosimetric evaluation tools, we also used the TCP and NTCPs for plan evaluation. In calculating TCP and NTCP, the heterogeneous biology distributions need to be taken into account. TCP is calculated using Eqs. (2) and (3) and NTCP is assessed using Lyman's model. The Kutcher-Burman effective-volume DVH reduction method⁴⁴ is extended to include the nonuniform functional unit density distribution using Eq. (8) when transforming a nonuniform dose distribution into a uniform irradiation of an effective partial volume. Model parameters from Burman *et al.*⁵³ are listed in Table I for the NTCP calculation.

E. Case studies

A prostate case with two different hypothetical distributions of radiobiological parameters is used to test the proposed BCRT inverse planning scheme. In each study, the target consists of the prostate gland with a few intraprostatic lesions. The sensitive structures include the rectum, bladder, and femoral heads. Figures 1(a) and 3(a) show the geometric shapes and locations of the structures in the two examples.

In the first example the target includes four biologically different regions, and the functional unit density distributions in the sensitive structures are uniform. Region 1 represents the basis reference target volume with typical parameters^{54,55} $\rho_{01} = 5 \times 10^5$ clonogen/cm³, $\alpha_1 = 0.26$ Gy⁻¹, and $\gamma_1 = \ln 2/40$ day⁻¹. The radiobiological parameters of the intra-

TABLE II. Radiological parameters for the target regions in the two examples.

Targets		ρ_{0i} (clonogen/cm ³)	α_i (Gy ⁻¹)	γ_i (day ⁻¹)
Example 1	Region 2	5×10^8	0.26	ln 2/40
	Region 3	5×10^5	0.13	ln 2/40
	Region 4	5×10^5	0.26	ln 2/10
Example 2	Region 2	5×10^6	0.20	ln 2/10
	Region 3	5×10^3	0.10	ln 2/60

prostatic lesions are listed in Table II. The parameters n_σ characterizing the dose-volume effect of the sensitive structures in the objective function (9) can be found in Table I. The parameter n_τ is chosen to be -0.2. For comparison, five IMRT plans, indexed by plan 1, -2, -3, -4, and -5, are generated. Plan 1 is obtained using the BCRT optimization scheme described above with $D_{\text{ref}}=70$ Gy. Plan 2 is obtained by prescribing the whole target a uniform dose of 70 Gy. Plan 3 and -4 are similar to plan 2 except that the dose is escalated to 81 and 91 Gy,^{12,14} respectively. Plan 5 is the SIB IMRT plan with the same prescribed doses as that of the BCRT. In plan 1 to -4, the objective function expressed in Eq. (9) is used and in plan 5 the conventional dose-based quadratic objective function is adopted. The optimization parameters (maximum dose constraints and importance factors) in the dose-based method were adjusted by trial and error to obtain the "optimal" plan. The same beam configuration (five equally spaced 15 MV photon beams with gantry angles of 0°, 72°, 144°, 216°, and 288° in IEC convention) is used in generating the five plans.

In the second example we hypothetically introduced a higher functional unit density region in the rectum [R_region 2 as shown in Fig. 3(a)] in addition to three biologically different target regions. The functional unit density of the R_region 1 is assigned a value of 1 and that of the R_region 2 is set to be 4. The same set ($\rho_{0i}, \alpha_i, \gamma_i$) as the previous example and a reference dose of 70 Gy are assigned to the prostate gland. The parameters for other target regions are listed in Table II. Once again, five IMRT plans are generated: Plan 1 is obtained using the proposed selective dose escalation scheme, plan 2, -3, and -4 are generated using different uniform prescription doses (70, 81, and 91 Gy) and plan 5 is SIB plan with the same prescription as plan 1 but is optimized using the conventional quadratic objective function. In generating these five plans, seven equally spaced 15 MV photon beams (0°, 51°, 103°, 154°, 206°, 257°, and 309°) are employed.

III. RESULTS

A. Example 1: Prostate case with four biologically different regions

In the first example, based on Eq. (7) and the parameters listed in Table II, the prescription doses to the target region 2, 3, and 4 are determined to be 85, 119, and 75 Gy, respectively. In order to examine the capability of the BCRT inverse planning system in producing an extremely nonuni-

form dose within a target volume, we have used an "extreme" combination of $\{\rho, \alpha, \gamma\}$, which leads to an exceedingly high prescription dose in region 3 (119 Gy). Figures 1(b)–1(d) show the isodose distributions of plan 1 in a transverse slice and two sagittal slices. The EDVH of the target and the DVHs of the sensitive structures are plotted in Fig. 2 for plan 1 in solid curves. For comparison, the corresponding EDVHs and DVHs of plan 2, -3, -4, and -5 are also shown in the figures as dashed, dotted, dash-dotted, and dash-dot-dotted curves, respectively. As seen from Fig. 1, all regions in the prostate are well covered by their prescription doses and the sensitive structures are well spared. Even in this extreme case, it seems that the inverse planning system can satisfy the biological requirement. A steep dose gradient is found at the interface between the target and the rectum. A comparison of the target EDVH in Fig. 2(a) indicates that above 98.5% of the target voxels achieved their desired doses in plan 1 and plan 5. However, for the uniform dose escalation scheme, the desired doses in some regions (region 2, 3 and part of region 4 in plan 2; region 2 and 3 in plan 3; and region 3 in plan 4) are not achieved. We found that, in plan 1, the doses to the surrounding sensitive structures are not significantly increased compared with those of plan 2, despite the fact that some voxels in region 4 receive a dose as high as 119 Gy. In plan 1, the rectum, bladder, and femoral heads are better spared in comparison with plan 3 and -4. However, by comparing the DVHs of plan 1 and -5, it is noticed that, although the target coverage in plan 5 is similar to that in plan 1, the sensitive structures in plan 5 receive much higher doses than plan 1, indicating that the proposed approach can improve the sensitive structure sparing compared with the conventional dose-based quadratic objective function. In addition, as can be expected, the target doses in plan 1 and -5 are less uniform in the target volume in comparison with that of plan 2, -3, and -4. This is more pronounced in the target region 1, where about 50% of the volume receives a dose larger than 85 Gy as shown in Fig. 2(b), resulting in an effective dose above 120% in ~50% of the target voxels [see Fig. 2(a)]. However, the increase of dose inhomogeneity is desirable here provided that the NTCPs are not compromised.

Table III lists the calculated TCPs for the targets and NTCPs for the sensitive structures with consideration of heterogeneous biology in all plans. It is seen that the overall TCPs for the three plans with uniform target dose prescriptions (plan 2, -3, and -4) are all less than that of the BCRT plan (plan 1) and SIB plan (plan 5). This is understandable because, in plan 2, -3 and -4, some target regions (such as target region 3) receive doses much less than the desired doses. For example, in plan 4, the TCP for target region 3 is only 0.461. Even if the TCPs for region 1, 2, and 4 are all close to 1.00, the resultant total TCP for plan 4 is 0.461. In contrast, the TCPs of plan 1 and plan 5 are 0.984 and 0.981, respectively. Furthermore, we found that the NTCPs of the sensitive structures in plan 1 are very close to plan 2, significantly less than plan 3, -4, and -5. For example, the rectum NTCPs are 0.21% for plan 1 and 0.20% for plan 2. These are

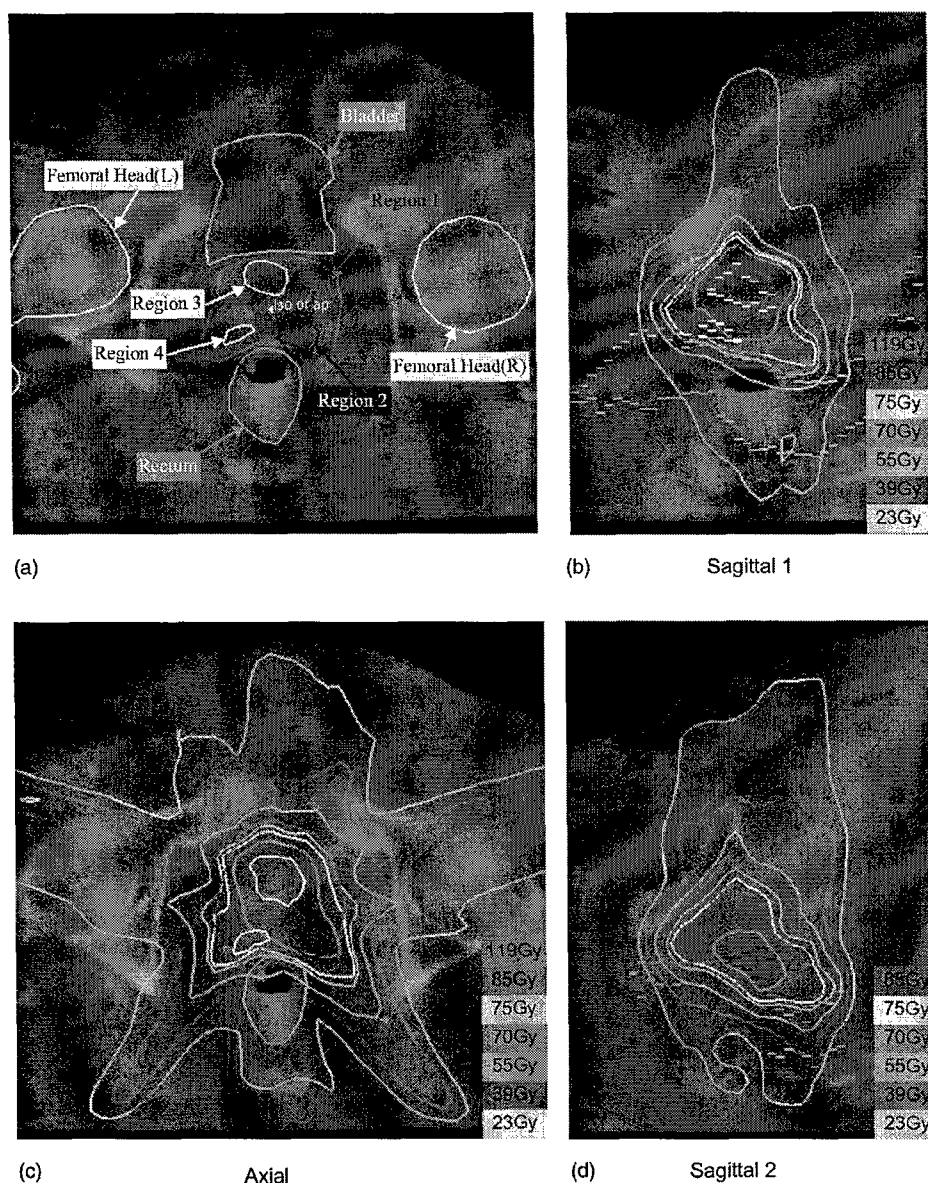


FIG. 1. A hypothetical prostate case with four biologically different regions (example 1). (a) Geometric shapes and locations of the targets and sensitive structures; (b)–(d): Isodose distributions in an axial slice and two sagittal slices for plan 1, generated by optimizing the objective function with a nonuniform dose prescription derived from Eq. (7).

increased to 0.65%, 1.84%, and 0.89% for Plan 3, -4, and -5, respectively. Again, although similar overall TCPs are achieved for the BCRT and dose-based SIB IMRT plans when the same dose prescriptions are used, the rectum NTCPs are significantly reduced using the proposed formalism. This is consistent with our previous study of the objective function in the context of conventional IMRT aiming to deliver a uniform dose to the target volume.⁴⁹

B. Example 2: Prostate case with three biologically different regions and nonuniform importance in rectum

In the second hypothetical example, there are three biologically different regions in the prostate and two functionally different regions in the rectum. The prescription doses for the three target regions are 70 (reference dose), 99, and 121 Gy, as determined by Eq. (7) with the biological param-

eters listed in Table II. Figures 3(b)–3(f) show the isodose distributions of plan 1 in three transverse slices and two sagittal slices. The EDVHs and DVHs of the target and sensitive structures for plan 1 to plan 5 are plotted in Fig. 4 as solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves, respectively. Similar to the previous example, in plan 1, all regions in the prostate are well covered by a dose comparable to the prescription and the sensitive structures are well spared. The dose gradient at the interface between the target and the rectum is very sharp for all the plans. From Fig. 4(a) we find that above 98% of the target voxels achieved their desired doses in plan 1. As a consequence of incorporating functional unit density information in inverse planning, the rectum sparing is even better than that of plan 2, much better than that of plan 3, -4. However, we notice that the sparing of the femoral heads in plan 1 is not as good as that in plan 2, -3, and -4. This is because high-intensity beamlets that pass

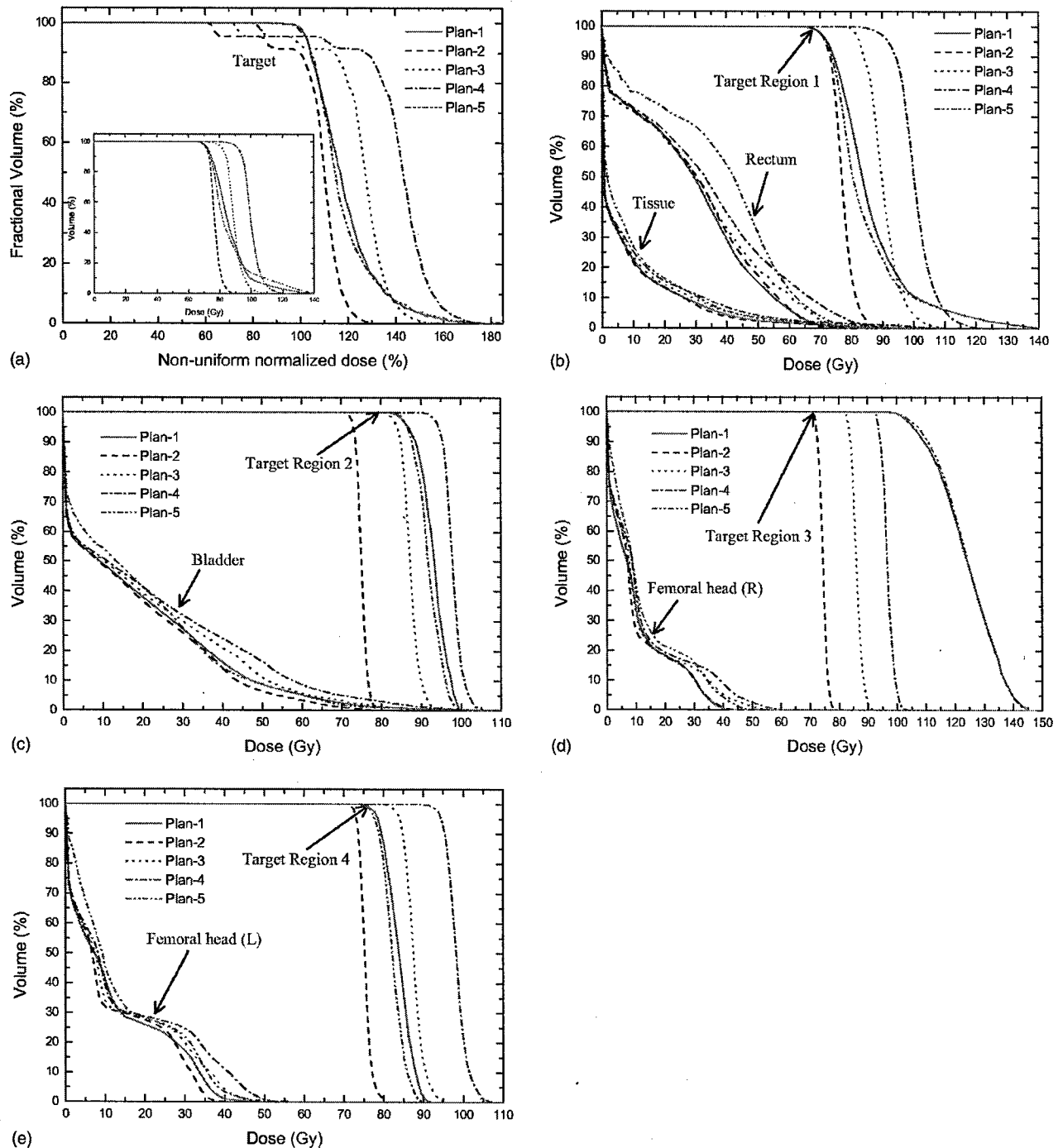


FIG. 2. Comparison of EDVHs and DVHs of the BCRT plan (plan 1), three uniform IMRT plans (plan 2: 70 Gy, plan 3: 81 Gy, and plan 4: 91 Gy) and the SIB plan (plan 5) in example 1. (a) Target EDVHs for the five plans (insert is the regular DVHs of the prostate target). The normalized doses to the target region 1, 2, 3 and 4 are 70, 85, 119, and 75 Gy, respectively; (b)-(e) DVHs of different target regions and sensitive structures for the five plans. The solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves represent the results of plan 1, 2, 3, 4, and 5, respectively. The effective dose is defined as the physical dose at a voxel normalized by its desired dose determined by Eq. (7).

through the femoral heads are needed to adequately irradiate the target region 3, as seen from Figs. 3(b) and 3(c). In addition, similar to the first example, the target coverage in plan 5 is close to that in plan 1, but the doses to the sensitive structures in plan 5 are much higher than that in plan 1.

Table IV lists the calculated TCPs and NTCPs for all plans. Once again, we found that the TCP of the target in the proposed BCRT technique is much higher and the NTCP of the rectum is lower compared with those obtained using the conventional uniform dose escalation schemes. Remarkably,

TABLE III. Comparison of TCP and NTCP for the five IMRT plans for example 1.

		Plan 1 (BCRT plan)	Plan 2 (70 Gy Uniform)	Plan 3 (81 Gy Uniform)	Plan 4 (91 Gy Uniform)	Plan 5 (SIB plan)
TCP	Region 1	0.997	0.995	1.000	1.000	0.994
	Region 2	0.998	0.642	0.995	1.000	0.999
	Region 3	0.989	0.000	0.002	0.461	0.989
	Region 4	1.000	0.997	1.000	1.000	0.998
	Overall	0.984	0.000	0.002	0.461	0.981
NTCP (%)	Rectum	0.212	0.196	0.652	1.84	0.885
	Bladder	1.6×10^{-5}	1.4×10^{-5}	2.3×10^{-5}	4.2×10^{-5}	3.6×10^{-5}
	Femoral head (R)	2.0×10^{-5}	2.1×10^{-6}	2.6×10^{-4}	1.75×10^{-4}	3.9×10^{-5}
	Femoral head (L)	1.2×10^{-5}	2.0×10^{-6}	7.0×10^{-4}	5.26×10^{-4}	6.9×10^{-5}

the overall TCP for the target is increased from 0.823 to 0.982 and the NTCP of the rectum is reduced from 3.1% to 0.40% when plan 4 is replaced by the selective dose escalation scheme (plan 1). Again, we found that, for similar overall TCPs, the rectum NTCPs of the BCRT plan are much lower in comparison with that obtained using dose-based SIB scheme.

IV. DISCUSSION

Equation (7) provides a general formula for determining the desired target dose distribution based on the known biology information of the system, and represents one of the main results of this study. A few special cases are worth discussing here. First, when the biology distribution is uniform in the target, a uniform dose of D_{ref} is desired. This is consistent with previous studies³⁷ and existing clinical knowledge.

When the clonogen density ρ is nonuniform while the values of α and γ are constant across the target, we have

$$D_0^T(i) = D_{\text{ref}} - \frac{1}{\alpha'} \ln \left(\frac{\rho_{0\text{ref}}}{\rho_{0i}} \right), \quad (10)$$

which is identical to the formula obtained by Webb and Nahum.³⁶ Equation (10) indicates that the desired dose depends on the tumor cell density logarithmically and is thus relatively insensitive to a variation in the clonogen density. For $\alpha_{\text{ref}}=0.312$, for example, even if the clonogen density in a tumor voxel is 10 times higher than that of the reference

situation, the desired dose is only about 7 Gy higher than the reference value. A detailed discussion of this special situation has been presented by Webb and Nahum.³⁶

Another special case is that the tumor clonogen density and the proliferation rate are constant and the radiosensitivity α is spatially nonuniform. Equation (7) now becomes

$$D_0^T(i) = \frac{\alpha'_{\text{ref}}}{\alpha'_i} D_{\text{ref}} - \frac{1}{\alpha'_i} \ln \left(\frac{\alpha'_{\text{ref}}}{\alpha'_i} \right). \quad (11)$$

The desired dose is approximately inversely proportional to the parameter α'_i and is thus sensitively dependent on the value of parameter α'_i . This is similar to the conclusions of Ebert and Hoban⁴⁰ and Levin-Plotnik and Hamilton.⁴¹ For example, if α'_i is reduced from 0.312 (corresponding to $\alpha=0.26$, fractional dose $d_i=2.0$ Gy, and α/β ratio=10 Gy) to 0.18 (corresponding to $\alpha=0.15$, $d_i=2.0$ Gy, and $\alpha/\beta=10$ Gy), the desired dose is increased by about 70% (from 70 Gy to about 118 Gy).

If we keep the tumor clonogen density and radiosensitivity α' constant and only allow the proliferation rate γ to vary spatially, then

$$D_0^T(i) = D_{\text{ref}} + \frac{1}{\alpha'_i} (\gamma_i - \gamma_{\text{ref}}) \Delta T. \quad (12)$$

Thus, the desired dose increases linearly with the proliferation rate. In this work the potential cell-doubling times, T_p , used by King *et al.*⁵⁴ are adopted. Since T_p for a prostate tumor is relatively longer, its influence on the desired dose is

TABLE IV. Comparison of TCP and NTCP for the five IMRT plans for example 2.

		Plan 1 (BCRT plan)	Plan 2 (70 Gy Uniform)	Plan 3 (81 Gy Uniform)	Plan 4 (91 Gy Uniform)	Plan 5 (SIB Plan)
TCP	Region 1	0.997	0.995	1.000	1.000	0.968
	Region 2	0.989	0.000	0.587	0.981	0.987
	Region 3	0.996	0.006	0.408	0.839	0.990
	Overall	0.981	0.000	0.239	0.823	0.946
NTCP (%)	Rectum	0.397	0.414	1.46	3.12	1.25
	Bladder	1.5×10^{-5}	1.2×10^{-5}	1.8×10^{-5}	4.3×10^{-5}	3.9×10^{-5}
	Femoral head (R)	3.7×10^{-5}	1.5×10^{-5}	1.8×10^{-5}	5.3×10^{-5}	2.3×10^{-5}
	Femoral head (L)	4.9×10^{-5}	1.1×10^{-5}	3.0×10^{-5}	4.5×10^{-5}	3.6×10^{-5}

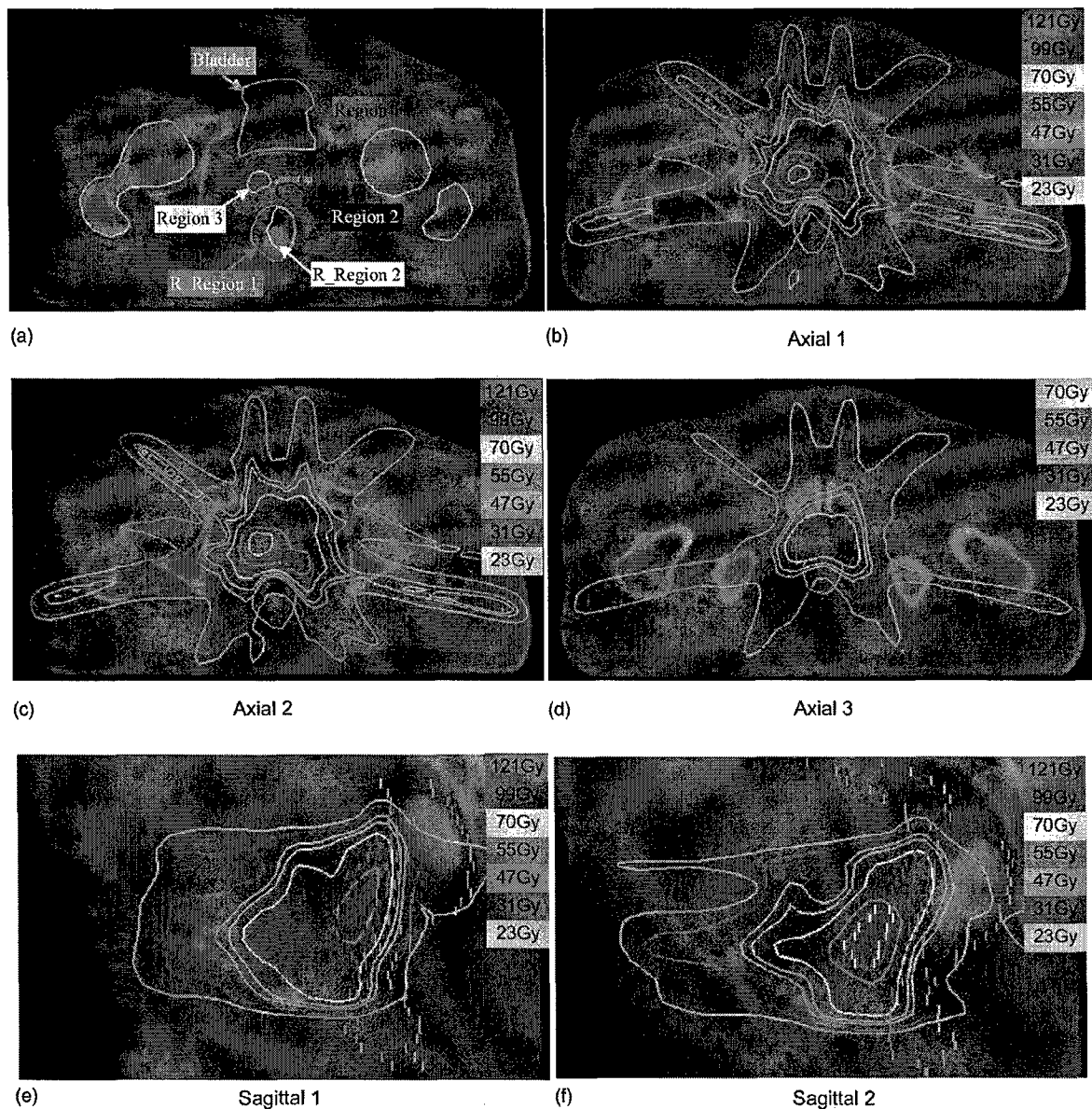


FIG. 3. A hypothetical prostate case with three biologically different regions and nonuniform importance in the rectum (example 2). (a) Geometric shapes and locations of the targets and sensitive structures; (b)–(d) Isodose distributions in three axial slices and two sagittal slices for plan 1, generated by optimizing the objective function with nonuniform dose prescription derived from Eq. (7).

not very significant. However, for other more rapidly proliferating tumors, the proliferation rate may play an important role. In such situations, reducing the overall treatment time ΔT (e.g., using an accelerated scheme) is helpful to minimize the influence of the proliferation rate.

We emphasize that the quadratic term in the linear-quadratic model plays an important role in accounting for the fractionation effect. If only the linear term is kept, the total dose $D_0^T(i)$ in Eq. (7) is no longer entangled with the fractional dose d_i . When the quadratic term is “switched on,” the value of $D_0^T(i)$ depends not only on the total reference dose but also on the fractional dose. For a large fractional dose, the total dose will be less, and vice versa. In other words, the total dose received by a voxel is determined by two contrib-

uting factors, one being the local biological parameters $\{\rho, \alpha, \gamma\}$, and the other being the coupling between the fractional dose and the total dose. The latter is responsible for the phenomenon that the total dose needs to be decreased when the number of fractions is reduced. If the quadratic term were ignored, according to Eq. (7), the dose required at a voxel would be much higher. For example, the desired doses for target region 3 in example 1 are determined to be 119 and 135 Gy with and without inclusion of the quadratic term, respectively.

We also would like to emphasize that in this study, the radiosensitivity α' and proliferation rate γ are assumed to be constants during the whole treatment course. In reality, both α' and γ may change with time due to such biological pro-

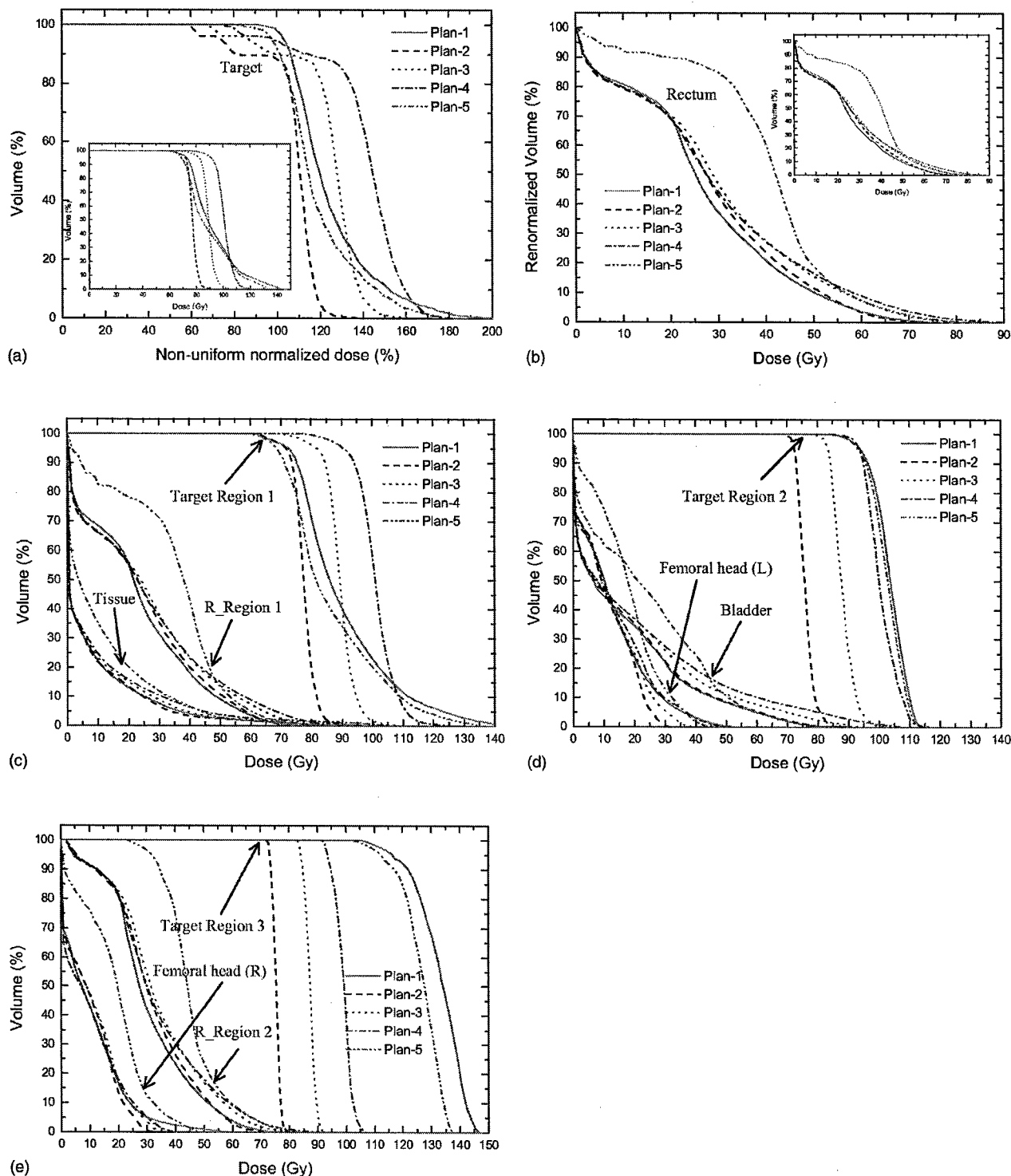


FIG. 4. Comparison of EDVHs, FDVHs, and DVHs of the BCRT plan (plan 1), three uniform IMRT plans (plan 2: 70 Gy, plan 3: 81 Gy, and plan 4: 91 Gy) and the SIB plan (plan 5) in example 2. (a) The target EDVHs for the five plans (insert is the regular DVHs of the prostate target). The normalized doses to the target region 1, 2, and 3 are 70, 99, and 121 Gy, respectively; (b) The rectum FDVHs for the five plans (insert is the regular DVHs of the rectum); (c)–(e): DVHs of different target regions and sensitive structures for the five plans. The solid, dashed, dotted, dash-dotted, and dash-dot-dotted curves represent the results of plan 1, 2, 3, 4, and 5, respectively.

cesses as tumor cell redistribution⁵⁶ and reoxygenation.⁵⁷ The time dependence of these factors may result in a reduction of the desired prescription dose, and this effect should be investigated in the future.

Comparing with the uniform dose escalation scheme, our study clearly suggests that deliberately incorporating an inhomogeneous dose distribution significantly enhances the TCP and reduces the NTCP. Physically, we believe that the

significant improvement arises from the more effective use of radiation in the newly proposed treatment scheme. A great deal of dose is "wasted" in the conventional uniform dose escalation scheme. For example, in the first example the increased doses in the target region 1 and 4 have almost no contributions to the enhancement of the TCP when plan 2 (70 Gy uniform dose to the prostate gland) is replaced by plan 3 (81 Gy) or plan 4 (91 Gy). Even though part of the prostate receives high doses in the selective dose escalation scheme (for example, 119 Gy in target region 3 of the first example), the total deposited energy in the targets is still less than that of plan 3 or -4. It is thus not difficult to understand why deliberately nonuniform dose distributions can, in general, reduce the radiation side effects and represent a more intelligent way to irradiate the tumor target.

A similar deficiency also exists in the current SIB approach. Although it is clear that the regions with different tumor burdens should be given different doses, the specific values for the regions are determined in an *ad hoc* manner. The empirical boost dose could be too low, in which case the tumor control is sacrificed, or too high, in which case other parts of the system are compromised. The problem is aggravated when the tumor burden varies continuously from point to point. In the proposed BCRT approach, the prescribed dose is voxel dependent and is determined based on the tumor biology distribution. In addition, a more sophisticated objective function is developed to take the dose-volume effect and functional density information of the sensitive structures into account, resulting in better sparing of the sensitive structures.

Finally, it should be recognized that our knowledge of radiobiological parameters for tumors or normal tissues is still very crude and the validity of the model is still under establishment. Therefore, the LQ model and the parameters adopted in the paper are fine for a proof of principle but they should not be taken as more than that.

V. CONCLUSION

In the presence of nonuniform biology distributions, IMRT inverse planning is complicated by the fact that it is not clear what represents the appropriate spatial dose prescription, which is generally used as a landmark to guide the dose optimization process. In this work, we have described a technique for deriving the prescription dose based on an LQ model with consideration of the cell proliferation. The relation is quite general and can be used as prescription dose to guide an arbitrary inverse planning objective function aimed to produce customized dose distribution in accordance with the spatial biology information. For a given patient, IMRT inverse planning now consists of two steps: Derivation of the prescription dose, and beam profile optimization that produces as closely as possible this prescription dose. The formalism proposed here lays a technical foundation for future BCRT development, allowing us to escalate tumor dose more intelligently and effectively. When combined with state-of-the-art biological imaging techniques, which promise to re-

veal detailed patient-specific biology distribution information, this study may have significant implication for the management of cancer in the future.

ACKNOWLEDGMENTS

The authors wish to thank Dr. B. Thorndyke for carefully reading the manuscript. We would also like to thank Dr. S. Hancock, Dr. A. Koong, Dr. C. King, Dr. Q. T. Le, Dr. A. Boyer, and Dr. D. Spielman for useful discussion. This research is supported in part by the National Cancer Institute (5 R01 CA98523-01) and Department of Defense (DAMD17-03-1-0019 and 0023).

APPENDIX: DERIVATION OF EQUATION (7)

We present the detailed derivation process for Eq. (7) under the condition of equal mass and volume for all target voxels.

Substituting Eq. (5) into Eq. (6), we obtain

$$\lambda \text{TCP}_i \frac{\partial(m_i D_i)}{\partial \text{TCP}_i} = \lambda \text{TCP}_{\text{ref}} \frac{\partial(m_{\text{ref}} D_{\text{ref}})}{\partial \text{TCP}_{\text{ref}}} = -\text{TCP}. \quad (\text{A1})$$

Since $\lambda \neq 0$ otherwise, TCP becomes zero according to the requirement of Eq. (6), which corresponds to a minimum. If we assumed that mass for all target voxels is equal, then Eq. (A1) becomes

$$\text{TCP}_i \frac{\partial D_i}{\partial \text{TCP}_i} = \text{TCP}_{\text{ref}} \frac{\partial D_{\text{ref}}}{\partial \text{TCP}_{\text{ref}}}. \quad (\text{A2})$$

From Eq. (2) we have

$$D_i = \frac{-1}{\alpha'_i} \ln \left\{ -\gamma_i \Delta T + \ln \left[\frac{-1}{\rho_i V_i} \ln \text{TCP}_i \right] \right\}. \quad (\text{A3})$$

Substituting the expressions from Eq. (A3) for both D_i and D_{ref} into Eq. (A2), we have

$$\alpha'_i \ln(\text{TCP}_i) = \alpha'_{\text{ref}} \ln(\text{TCP}_{\text{ref}}). \quad (\text{A4})$$

The desired doses, $D_0^T(i)$, producing maximum TCP with the constraint of constant integral dose, can be obtained by substituting TCP_i and TCP_{ref} expressed in Eq. (2) into Eq. (A4)

$$D_0^T(i) = \frac{\alpha'_{\text{ref}}}{\alpha'_i} D_{\text{ref}} - \frac{1}{\alpha'_i} (\gamma_{\text{ref}} - \gamma_i) \Delta T - \frac{1}{\alpha'_i} \ln \left(\frac{\alpha'_{\text{ref}} \rho_{\text{ref}} V_{\text{ref}}}{\alpha'_i \rho_i V_i} \right). \quad (\text{A5})$$

When volume for all target voxels is equal, Eq. (A5) becomes Eq. (7).

^{a)} Author to whom correspondence should be addressed. Electronic mail: lei@reyes.stanford.edu

¹ S. Webb, *Intensity-Modulated Radiation Therapy* (Institute of Physics, Bristol, 2001), Vol. XV.

² C. C. Ling, C. Burman, C. S. Chui, G. J. Kutcher, S. A. Leibel, T. LoSasso, R. Mohan, T. Bortfeld, L. Reinstein, S. Spirou, X. H. Wang, Q. Wu, M. Zelefsky, and Z. Fuks, "Conformal radiation treatment of prostate cancer using inversely-planned intensity-modulated photon beams produced with dynamic multileaf collimation," *Int. J. Radiat. Oncol., Biol., Phys.* **35**, 721-730 (1996).

³ S. Y. Woo, W. H. Grant 3rd, D. Bellezza, R. Grossman, P. Gildenberg, L.

- S. Carpentar, M. Carol, and E. B. Butler, "A comparison of intensity modulated conformal therapy with a conventional external beam stereotactic radiosurgery system for the treatment of single and multiple intracranial lesions," *Int. J. Radiat. Oncol., Biol., Phys.* **35**, 593–597 (1996).
- ⁴N. Lee, P. Xia, J. M. Quivey, K. Sultanem, I. Poon, C. Akazawa, P. Akazawa, V. Weinberg, and K. K. Fu, "Intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: An update of the UCSF experience," *Int. J. Radiat. Oncol., Biol., Phys.* **53**, 12–22 (2002).
- ⁵Q. T. Le, L. Xing, and A. L. Boyer, "Head and Neck IMRT," in *3D Conformal Radiation Therapy & Intensity Modulated Radiation Therapy in the Next Millennium*, edited by J. G. Purdy and J. Palta (Medical Physics Publishing, Madison, WI, 1999), pp. 190–195.
- ⁶E. Huang, B. S. Teh, D. R. Strother, Q. G. Davis, J. K. Chiu, H. H. Lu, L. S. Carpenter, W. Y. Mai, M. M. Chintagumpala, M. South, W. H. Grant 3rd, E. B. Butler, and S. Y. Woo, "Intensity-modulated radiation therapy for pediatric medulloblastoma: Early report on the reduction of ototoxicity," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 599–605 (2002).
- ⁷J. G. Rosenman, J. S. Halle, M. A. Socinski, K. Deschesne, D. T. Moore, H. Johnson, R. Fraser, and D. E. Morris, "High-dose conformal radiotherapy for treatment of stage IIIA/IIIB non-small-cell lung cancer: Technical issues and results of a phase I/II trial," *Int. J. Radiat. Oncol., Biol., Phys.* **54**, 348–356 (2002).
- ⁸A. J. Mundt, A. E. Lujan, J. Rotmensch, S. E. Waggoner, S. D. Yamada, G. Fleming, and J. C. Roeske, "Intensity-modulated whole pelvic radiotherapy in women with gynecologic malignancies," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 1330–1337 (2002).
- ⁹M. J. Krasin, B. T. Crawford, Y. Zhu, E. S. Evans, M. R. Sontag, L. E. Kun, and T. E. Merchant, "Intensity-modulated radiation therapy for children with intraocular retinoblastoma: Potential sparing of the bony orbit," *Clin. Oncol. (R. Coll. Radiol)* **16**, 215–222 (2004).
- ¹⁰J. Meyer, M. H. Phillips, P. S. Cho, I. Kalet, and J. N. Doctor, "Application of influence diagrams to prostate intensity-modulated radiation therapy plan selection," *Phys. Med. Biol.* **49**, 1637–1653 (2004).
- ¹¹A. Eisbruch, "Intensity-modulated radiotherapy of head-and-neck cancer: Encouraging early results," *Int. J. Radiat. Oncol., Biol., Phys.* **53**, 1–3 (2002).
- ¹²M. J. Zelefsky, Z. Fuks, M. Hunt, Y. Yamada, C. Marion, C. C. Ling, H. Amols, E. S. Venkatraman, and S. A. Leibel, "High-dose intensity modulated radiation therapy for prostate cancer: Early toxicity and biochemical outcome in 772 patients," *Int. J. Radiat. Oncol., Biol., Phys.* **53**, 1111–1116 (2002).
- ¹³S. M. Zhou, L. B. Marks, G. S. Tracton, G. S. Sibley, K. L. Light, P. D. Maguire, and M. S. Anscher, "A new three-dimensional dose distribution reduction scheme for tubular organs," *Med. Phys.* **27**, 1727–1731 (2000).
- ¹⁴C. C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J. Koutcher, "Towards multidimensional radiotherapy (MD-CRT): Biological imaging and biological conformality," *Int. J. Radiat. Oncol., Biol., Phys.* **47**, 551–560 (2000).
- ¹⁵A. Brahme, "Individualizing cancer treatment: Biological optimization models in treatment planning and delivery," *Int. J. Radiat. Oncol., Biol., Phys.* **49**, 327–337 (2001).
- ¹⁶L. Xing, C. Cotrutz, S. Hunjan, A. L. Boyer, E. Adalsteinsson, and D. Spielman, "Inverse planning for functional image-guided intensity-modulated radiation therapy," *Phys. Med. Biol.* **47**, 3567–3578 (2002).
- ¹⁷K. S. Chao, S. Mutic, R. L. Gerber, C. A. Perez, and J. A. Purdy, "A novel approach to overcome hypoxic tumor resistance: Cu-ATSM-guided intensity-modulated radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **49**, 1171–1182 (2001).
- ¹⁸R. Macklis, M. Weinhaus, and G. Harnisch, "Intensity-modulated radiotherapy: Rethinking basic treatment planning paradigms," *Int. J. Radiat. Oncol., Biol., Phys.* **48**, 317–318 (2000).
- ¹⁹M. Alber, F. Paulsen, S. M. Eschmann, and H. J. Machulla, "On biologically conformal boost dose optimization," *Phys. Med. Biol.* **48**, N31–35 (2003).
- ²⁰Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, "Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 224–235 (2002).
- ²¹S. R. Cherry, "In vivo molecular and genomic imaging: New challenges for imaging physics," *Phys. Med. Biol.* **49**, R13–48 (2004).
- ²²J. Kurhanewicz, M. G. Swanson, S. J. Nelson, and D. B. Vigneron, "Combined magnetic resonance imaging and spectroscopic imaging approach to molecular imaging of prostate cancer," *J. Magn. Reson. Imaging* **16**, 451–463 (2002).
- ²³C. H. Contag and B. D. Ross, "It's not just about anatomy: In vivo bioluminescence imaging as an eyepiece into biology," *J. Magn. Reson. Imaging* **16**, 378–387 (2002).
- ²⁴S. J. DiBiase, K. Hosseinzadeh, R. P. Gullapalli, S. C. Jacobs, M. J. Naslund, G. N. Sklar, R. B. Alexander, and C. Yu, "Magnetic resonance spectroscopic imaging-guided brachytherapy for localized prostate cancer," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 429–438 (2002).
- ²⁵J. G. Tjuvajev *et al.*, "Imaging of brain tumor proliferative activity with iodine-131-iododeoxyuridine," *J. Nucl. Med.* **35**, 1407–1417 (1994).
- ²⁶A. F. Shields, J. R. Grierson, B. M. Dohmen, H. J. Machulla, J. C. Stayanoff, J. M. Lawhorn-Crews, J. E. Obradovich, O. Muzik, and T. J. Mangner, "Imaging proliferation in vivo with [F-18] FLT and positron emission tomography," *Nat. Med.* **4**, 1334–1336 (1998).
- ²⁷J. S. Rasey, W. J. Koh, M. L. Evans, L. M. Peterson, T. K. Lewellen, M. M. Graham, and K. A. Krohn, "Quantifying regional hypoxia in human tumors with positron emission tomography of [18F]fluoromisonidazole: A pretherapy study of 37 patients," *Int. J. Radiat. Oncol., Biol., Phys.* **36**, 417–428 (1996).
- ²⁸W. J. Koh *et al.*, "Evaluation of oxygenation status during fractionated radiotherapy in human nonsmall cell lung cancers using [F-18]fluoromisonidazole positron emission tomography," *Int. J. Radiat. Oncol., Biol., Phys.* **33**, 391–398 (1995).
- ²⁹B. L. Miller, L. Chang, R. Booth, T. Ernst, M. Cornford, D. Nikas, D. McBride, and D. J. Jenden, "In vivo 1H MRS choline: Correlation with in vitro chemistry/histology," *Life Sci.* **58**, 1929–1935 (1996).
- ³⁰R. K. Gupta, T. F. Cloughesy, U. Sinha, J. Garakian, J. Lazareff, G. Rubino, L. Rubino, D. P. Becker, H. V. Vinters, and J. R. Alger, "Relationships between holine magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma," *J. Neuro-Oncol.* **50**, 215–226 (2000).
- ³¹R. J. Hamilton, M. J. Blend, C. A. Pelizzari, B. D. Milliken, and S. Vijayakumar, "Using vascular structure for CT-SPECT registration in the pelvis," *J. Nucl. Med.* **40**, 347–351 (1999).
- ³²A. F. Shields, D. A. Mankoff, J. M. Link, M. M. Graham, J. F. Eary, S. M. Kozawa, M. Zheng, B. Lewellen, T. K. Lewellen, J. R. Grierson, and K. A. Krohn, "Carbon-11-thymidine and FDG to measure therapy response," *J. Nucl. Med.* **39**, 1757–1762 (1998).
- ³³L. B. Marks, G. W. Sherouse, M. T. Munley, G. C. Bentel, D. P. Spencer, and C. Scarfone, "Incorporation of functional status into dose-volume analysis," *Med. Phys.* **26**, 196–199 (1999).
- ³⁴T. E. Schultheiss, G. K. Zagars, and L. J. Peters, "An explanatory hypothesis for early- and late-effect parameter values in the LQ model," *Radiation Oncol.* **9**, 241–248 (1987).
- ³⁵J. F. Fowler, "The linear-quadratic formula and progress in fractionated radiotherapy," *Br. J. Radiol.* **62**, 679–694 (1989).
- ³⁶S. Webb and A. E. Nahum, "A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density," *Phys. Med. Biol.* **38**, 653–666 (1993).
- ³⁷S. Webb, P. M. Evans, W. Swindell, and D. J. O., "A proof that uniform dose gives the greatest TCP for fixed integral dose in the planning target volume," *Phys. Med. Biol.* **39**, 2091–2098 (1994).
- ³⁸A. Brahme and A. K. Agren, "Optimal dose distribution for eradication of heterogeneous tumours," *Acta Oncol.* **26**, 377–385 (1987).
- ³⁹N. A. Stavreva, P. V. Stavrev, and W. H. Round, "A mathematical approach to optimizing the radiation dose distribution in heterogeneous tumours," *Acta Oncol.* **35**, 727–732 (1996).
- ⁴⁰M. A. Ebert and P. W. Hoban, "Some characteristics of tumour control probability for heterogeneous tumours," *Phys. Med. Biol.* **41**, 2125–2133 (1996).
- ⁴¹D. Levin-Plotnik and R. J. Hamilton, "Optimization of tumour control probability for heterogeneous tumours in fractionated radiotherapy treatment protocols," *Phys. Med. Biol.* **49**, 407–424 (2004).
- ⁴²T. E. Schultheiss, C. G. Orton, and R. A. Peck, "Models in radiotherapy: Volume effects," *Med. Phys.* **10**, 410–415 (1983).
- ⁴³J. T. Lyman and A. B. Wolbarst, "Optimization of radiation therapy. III. A method of assessing complication probabilities from dose-volume histograms," *Int. J. Radiat. Oncol., Biol., Phys.* **13**, 103–109 (1987).
- ⁴⁴G. J. Kutcher and C. Burman, "Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method," *Int. J. Radiat. Oncol., Biol., Phys.* **16**, 1623–1630 (1989).

- ⁴⁵R. Mohan, G. S. Mageras, B. Baldwin, L. J. Brewster, G. J. Kutcher, S. Leibel, C. M. Burman, C. C. Ling, and Z. Fuks, "Clinically relevant optimization of 3-D conformal treatments," *Med. Phys.* **19**, 933-944 (1992).
- ⁴⁶J. O. Deasy, K. S. Chao, and J. Markman, "Uncertainties in model-based outcome predictions for treatment planning," *Int. J. Radiat. Oncol., Biol., Phys.* **51**, 1389-1399 (2001).
- ⁴⁷M. K. Martel, W. M. Sahijdak, R. K. Ten Haken, M. L. Kessler, and A. T. Turrisi, "Fraction size and dose parameters related to the incidence of pericardial effusions," *Int. J. Radiat. Oncol., Biol., Phys.* **40**, 155-161 (1998).
- ⁴⁸L. Jones and P. Hoban, "A comparison of physically and radiobiologically based optimization for IMRT," *Med. Phys.* **29**, 1447-1455 (2002).
- ⁴⁹Y. Yang and L. Xing, "Clinical knowledge-based inverse treatment planning," *Phys. Med. Biol.* **49**, 5101-5117 (2004).
- ⁵⁰L. Xing and G. T. Y. Chen, "Iterative algorithms for inverse treatment planning," *Phys. Med. Biol.* **41**, 2107-2123 (1996).
- ⁵¹L. Xing, R. J. Hamilton, D. Spelbring, C. A. Pelizzari, G. T. Chen, and A. L. Boyer, "Fast iterative algorithms for three-dimensional inverse treatment planning," *Med. Phys.* **25**, 1845-1849 (1998).
- ⁵²Y. Lu, D. R. Spelbring, and G. T. Chen, "Functional dose-volume histograms for functionally heterogeneous normal organs," *Phys. Med. Biol.* **42**, 345-356 (1997).
- ⁵³C. Burman, G. J. Kutcher, B. Emami, and M. Goitein, "Fitting of normal tissue tolerance data to an analytic function," *Int. J. Radiat. Oncol., Biol., Phys.* **21**, 123-135 (1991).
- ⁵⁴C. R. King, T. A. DiPetrillo, and D. E. Wazer, "Optimal radiotherapy for prostate cancer: Predictions for conventional external beam, IMRT, and brachytherapy from radiobiologic models," *Int. J. Radiat. Oncol., Biol., Phys.* **46**, 165-172 (2000).
- ⁵⁵A. E. Nahum, B. Movsas, E. M. Horwitz, C. C. Stobbe, and J. D. Chapman, "Incorporating clinical measurements of hypoxia into tumor local control modeling of prostate cancer: Implications for the alpha/beta ratio," *Int. J. Radiat. Oncol., Biol., Phys.* **57**, 391-401 (2003).
- ⁵⁶S. L. Scott, P. H. Gumerlock, L. Beckett, Y. Li, and Z. Goldberg, "Survival and cell cycle kinetics of human prostate cancer cell lines after single- and multifraction exposures to ionizing radiation," *Int. J. Radiat. Oncol., Biol., Phys.* **59**, 219-227 (2004).
- ⁵⁷R. F. Kallman and M. J. Dorie, "Tumor oxygenation and reoxygenation during radiation therapy: Their importance in predicting tumor response," *Int. J. Radiat. Oncol., Biol., Phys.* **12**, 681-685 (1986).

Inverse treatment planning with adaptively evolving voxel-dependent penalty scheme

Yong Yang and Lei Xing

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305-5847

(Received 26 February 2004; revised 24 May 2004; accepted for publication 9 August 2004; published 21 September 2004)

In current inverse planning algorithms it is common to treat all voxels within a target or sensitive structure equally and use structure specific prescriptions and weighting factors as system parameters. In reality, the voxels within a structure are not identical in complying with their dosimetric goals and there exists strong intrastructural competition. Inverse planning objective function should not only balance the competing objectives of different structures but also that of the individual voxels in various structures. In this work we propose to model the intrastructural tradeoff through the modulation of voxel-dependent importance factors and deal with the challenging problem of how to obtain a sensible set of importance factors with a manageable amount of computing. Instead of letting the values of voxel-dependent importance to vary freely during the search process, an adaptive algorithm, in which the importance factors were tied to the local radiation doses through a heuristically constructed relation, was developed. It is shown that the approach is quite general and the EUD-based optimization is a special case of the proposed framework. The new planning tool was applied to study a hypothetical phantom case and a prostate case. Comparison of the results with that obtained using conventional inverse planning technique with structure specific importance factors indicated that the dose distributions from the conventional inverse planning are at best suboptimal and can be significantly improved with the help of the proposed nonuniform penalty scheme. © 2004 American Association of Physicists in Medicine.

[DOI: 10.1118/1.1799311]

Key words: IMRT, inverse planning, dose optimization, objective function, tradeoff

INTRODUCTION

Inverse treatment planning is an indispensable step in the implementation of IMRT and its performance critically determines the success of an IMRT treatment.¹⁻¹¹ While the general concept of inverse planning seems to be logical and straightforward, much remains to be done to improve the currently available algorithms.^{3,12-14} One of the key issues that have been overlooked in the formulation of the inverse planning problem is the intrastructural tradeoff. Indeed, in most, if not all, inverse planning algorithms, the dose prescription and weighting factors are specified on a structure specific basis. An implicit assumption made in these implementations is that all points within a structure are equivalent. In reality, voxels within a structure are generally not equivalent in complying with their dosimetric requirements and this inherent heterogeneity has not been considered so far. Depending on the patient's geometry, beam modality and field configuration, some regions in a patient may have better chance to meet the prescription than others. Furthermore, the doses at different regions are often incompatible and even compete each other. Appropriate tradeoff between the voxels is a requisition to fully exploit the potential of IMRT in any type of inverse planning techniques.¹⁵⁻¹⁹

In this work we propose to model the intrastructural tradeoff through the modulation of voxel-dependent importance factors. The new inverse planning technique allows us to balance not only the competing objectives of different

structures but also that of the individual voxels within any structure. Computationally, the determination of intrastructural tradeoff is an intensive task because of the coupling between the beamlet weights and local importance. To obtain an adequate set of local importance factors with a manageable amount of computing time, we develop an adaptive algorithm. The main idea of the adaptive approach is to heuristically link the values of the voxel-dependent importance factors with the corresponding radiation doses and to continuously update their values until the optimality criterion is met. The new planning technique has been applied to study a hypothetical phantom case and a prostate case and its advantage over the conventional inverse planning technique with structurally uniform importance has been demonstrated. The algorithm seems to afford an effective way of modeling the system with only a little computational overhead and shows significant promise to improve the existing inverse planning techniques.

METHODS AND MATERIALS

Theoretical background

Clinical objectives are multifaceted and often incompatible with one another. The inverse planning is thus inherently a multiobjective problem: each anatomical structure, or more generally, each voxel within a structure, as an independent entity with its own treatment objective. A commonly used

approach is to combine these objectives to form an overall objective function by weighted sum of the individual objectives. That is

$$F = \sum_{\sigma=1} r_{\sigma} F_{\sigma}(\{D_c(i)\}), \quad (1)$$

where the summation is over all targets and sensitive structures, $D_c(i)$ is the calculated dose in voxel i , and r_{σ} and F_{σ} are the weighting factor (or the importance factor) and the objective of the structure indexed by σ , respectively. Regardless of the model used, F_{σ} can be expressed as a function of the dose distribution, $\{D_c(i)\}$, within the structure. For a dose-based optimization, the quadratic form,²⁰⁻²²

$$F_{\sigma} = \frac{1}{N_{\sigma}} \sum_{i=1}^{N_{\sigma}} r_i [D_c(i) - D_0(i)]^2, \quad (2)$$

is often used, where N_{σ} represents the total number of voxels in structure σ , and $D_0(i)$ is the prescription dose in voxel i . In the above description, two types of weighting factors are involved: the interstructural importance, $\{r_{\sigma}\}$, and the intrastructural ones, $\{r_i\}$. Up to this point, however, only structure specific importance factors are utilized in inverse planning formulation and the importance of voxel specific weighting factors has been overlooked other than the fact that they have been used to "tweak" IMRT dose distributions.^{15,16,18} Indeed, in most, if not all, objective functions, the voxels within a structure are tacitly assumed to be equivalent in complying with their dosimetric requirements and their relative importance factors are set to unity. This type of penalty scheme seriously limits the solution space and often leads to suboptimal plan. To give a comprehensive example, one can imagine the consequence when two or more structures in a system are restricted to take a single importance. To be able to assess more candidate plans and obtain truly optimal IMRT plans, it is necessary to establish a voxel-dependent penalty scheme in which the penalty at a voxel depends not only on the dose deviation from the prescription but also on other physical and clinical requirements at the point, which can be reflected by purposely modulating the $\{r_i\}$.

Voxel-dependent importance factor and its heuristic relation to the local dose

While the introduction of voxel-dependent importance factors has the advantage for us to obtain IMRT solutions that would be otherwise not accessible, a practical question is how to determine the optimal importance distribution for a given structure. In reality, there may be more than one source that contribute to the voxel heterogeneity (for example, the in-equivalence of the voxels in complying their dosimetric goals as discussed in this work, or biological heterogeneity resulting from nonuniform cell density or radiation sensitivity distribution) and the local importance can be generally written as a product of the contributions from various sources, that is, $r_i = r_i^1 r_i^2 r_i^3 \dots$, where the superscripts index the different sources of contributions. A possible approach for obtaining the importance distribution is the so-called *a priori* technique, in which one attempts to identify the

origin of voxel in-equivalence and then derives the spatial distribution of $\{r_i\}$ based on physical or clinical considerations. One can also proceed in a *posteriori* fashion¹⁵⁻¹⁷ by directly optimizing the $\{r_i\}$, similar to the autoselection of the structure specific importance factors proposed by Xing *et al.*¹⁷ An optimization of the intrastructural tradeoff is likely to be computationally prohibitive due to the enormous size of the search space arising from the coupling of the beamlet weights and the local importance factor. To circumvent the problem, in this work we develop an alternative technique in which the importance distribution is determined adaptively under the guidance of a heuristically constructed function.

Intuitively, it is not difficult to conceive that the r_i in Eq. (2) can be regarded as a dose-dependent parameter. For a voxel in a sensitive structure, for instance, the value of r_i should be higher if the voxel receives a high dose so that the voxel gets more penalty, and *vice versa*. This suggests that the local importance can be expressed as a function of the local dose and updated at each step of the iterative dose optimization process without invoking any additional computation. While the general monotonic dependence of r_i on the dose is clear, its specific form is a matter of experimentation. We propose

$$r_i = a_0 + a_1 D_c(i)^k \quad (3)$$

for the application, where a_0 , a_1 , and k are heuristic structure specific parameters. The value of k is greater than zero for a voxel in a sensitive structure. For a voxel in a target, k should be less than zero in order to assimilate the clinical preference over overdosing versus underdosing. The penalty function for a structure consists essentially of two terms. The first term is the conventional one with uniform importance across the whole structure and the second term modulates the importance according to the local value of dose. This phenomenological importance emphasizes on penalizing those voxels receiving a high radiation dose yet does not ignore those voxels with intermediate and even low dose of radiation at each step of the iterative calculation. In this study, we set a_0 as unity and a_1 as $D_{\text{ref}}^{-k}(i)$, where $D_{\text{ref}}(i)$ is set as the 5% tolerance dose $TD_{5/5}$ for sensitive structures and prescription dose $D_0(i)$ for target. k is an adjustable parameter determined by a trial-and-error process. We emphasize that the $\{r_i\}$ distribution so obtained may not be the best possible distribution due to the restriction on the feasible values imposed by Eq. (3). However, as will be seen in the Results section, the nonuniform importance factors resulted from the use of Eq. (3) allow us to significantly improve the IMRT dose distributions obtained using the existing inverse planning algorithm based on structure specific importance factors. The technique thus provides a practical solution to what appears to be a computationally overwhelming problem.

Calculation process

An independent optimization module based on the objective function (2) [with $\{r_i\}$ determined by Eq. (3)] was integrated into the PLUNC treatment planning system (University of North Carolina, Chapel Hill, NC). The dose

calculation engine and varieties of evaluation tools existing in the PLUNC system were employed for this study. For a given patient, a set of structure specific importance factors, $\{r_o\}$, were determined empirically, similar to that in the conventional inverse planning. For a given set of $\{r_o\}$, we used the ray-by-ray iterative algorithm (SIITP) reported in Refs. 23 and 24 to obtain the optimal beam intensity profiles. The only difference here is that a r_i -modulated quadratic objective function was used. The calculation time for a typical prostate case is less than 5 minutes on a PC with P4 1.7 GHz and 1024 MB RAM.

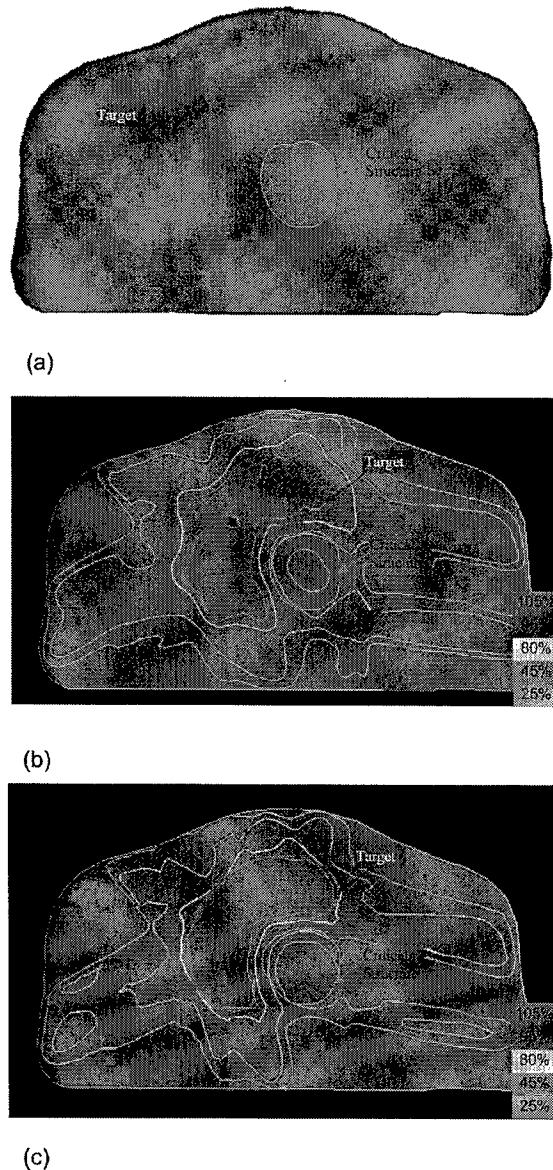


FIG. 1. The comparison of the isodose distributions for the hypothetical phantom case: (a) the geometry of the phantom; (b) dose distribution obtained using structurally uniform importance factors; and (c) dose distribution obtained using the newly proposed algorithm with structurally nonuniform importance factors.

TABLE I. Summary of the optimization parameters used in the conventional and newly proposed approaches for the hypothetical phantom case.

Organs	The dose-based approach		The proposed approach		
	Importance factors (r_o)	Prescription or tolerance doses (%)	Importance factors (r_o)	D_{ref} (%)	k
Target	8.0	100	15.0	100	-4
Critical Structure	5.0	35	0.6	45	10
Normal tissue	1.0	70	0.5	70	2

Case studies

A hypothetical phantom case and a previously treated clinical prostate case were used to evaluate the level of improvement resulted from the new algorithm with adaptively determined spatially nonuniform importance factors. The results for both cases were compared with that obtained using the conventional dose optimization method with structural specific importance factors. For the purpose of plan comparison and assessment, we purposely adjusted the structure specific importance factors in such a way that the dose coverage of the target volume was similar for the two IMRT plans.

The phantom case consisted of a target and a sensitive structure [Fig. 1(a)]. Five incident beams were used for target irradiation (32° , 104° , 176° , 248° and 320° in IEC convention). The incident photon energy is 15 MV. Table I listed the optimization parameters for both the approaches for this case. The target was prescribed to 100 (arbitrary unit) and the prescription dose to the sensitive structure, $D_0(i)$ in Eq. (2), was set to zero.

In the prostate IMRT case, the target volume was the prostate. The sensitive structures involved in this study were rectum and bladder. Five equally spaced 15 MV photon beams (gantry angles of 0° , 72° , 144° , 216° , and 288° in IEC convention) were used for the treatment. A radiation dose of 70 Gy was prescribed to cover 99% of the target volume. For the sensitive structures, the prescription doses, $D_0(i)$ in Eq. (2), were all set to zero. The optimization parameters for both approaches were summarized in Table II.

TABLE II. Summary of the optimization parameters used in the conventional and newly proposed approaches for the prostate case.

Organs	The dose-based approach		The proposed approach		
	Importance factors (r_o)	Prescription or tolerance doses (Gy)	Importance factors (r_o)	D_{ref} (Gy)	k
Target	4.0	70	5.0	70	-5
Bladder	1.0	50	0.1	65	2
Rectum	1.0	45	0.3	60	8.3
Normal tissue	0.5	65	0.4	68	2

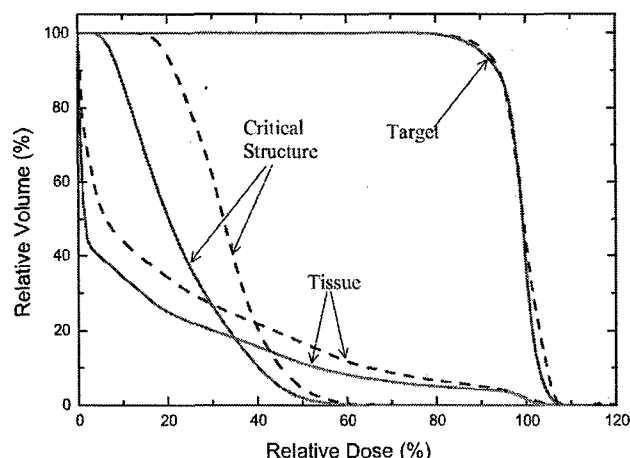


FIG. 2. Comparison of the dose volume histograms (DVH) for the hypothetical phantom IMRT plans obtained using our newly proposed algorithm (solid curves) and the conventional algorithm (dashed curves).

RESULTS AND DISCUSSION

Hypothetical phantom IMRT plans

Figures 1 and 2 summarized the calculation results and the comparison of the two IMRT plans obtained using the newly proposed and conventional penalty schemes. In Fig. 1(a) we depict the geometry of the phantom and in Figs. 1(b) and 1(c) we show the isodose distributions for the plans without and with the use of nonuniform intrastructural importance factors. The DVHs of the target and sensitive struc-

ture are plotted in Fig. 2. The dashed and solid curves in Fig. 2 represent the results of the conventional and the newly proposed inverse planning approaches, respectively. It is clearly seen that the proposed technique greatly improves the critical structure sparing. The maximum dose to the sensitive structure is reduced from 67 to 61 and the doses to other voxels in the sensitive structure are all reduced by 2%–14%. The fractional volume reduction in the intermediate and low doses is more distinct. For example, the fraction volume receiving dose above 40 is decreased from 19% to 10% while that receiving dose above 20 is decreased from 95% to 51%. For the target, the volume receiving dose above 85 is similar for the two algorithms, 98.4% and 98.6% for the conventional and the new approaches, respectively. However, the volume receiving high doses is decreased (e.g., the fractional volume receiving a dose above 102 is reduced from 20.7% to 8.2%). Interestingly, the irradiation to the normal tissue surrounding the target and sensitive structure seems to be less in the new optimization technique. For example, the fraction volumes of the normal tissue that received doses above 60 and 40 are decreased from 12% to 8% and from 34% to 25%, respectively. We attribute the simultaneous improvement in the doses of all structures to the greatly enlarged solution space when nonuniform importance is permissible.

Prostate IMRT plans

The IMRT plans obtained using the two different approaches for the prostate case are summarized in Figs. 3 and 4. Figure 3 compares the isodose distributions in two trans-

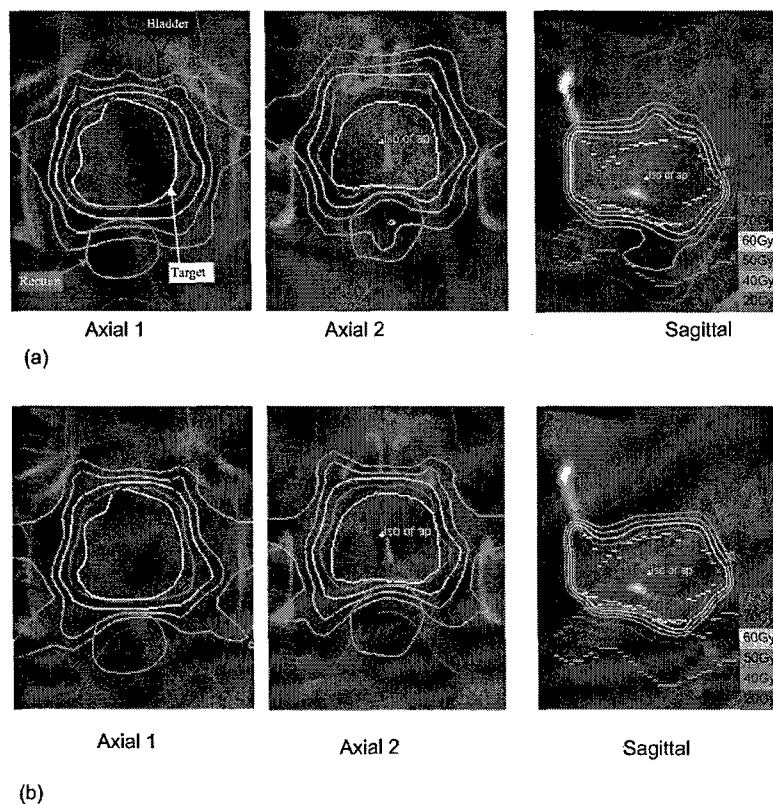


FIG. 3. Comparison of the isodose distributions for the prostate IMRT plans: (a) dose distribution obtained using structurally uniform importance factors; (b) dose distribution obtained using the newly proposed algorithm with structurally nonuniform importance factors. The results on two transverse slices, and one sagittal slice are shown.

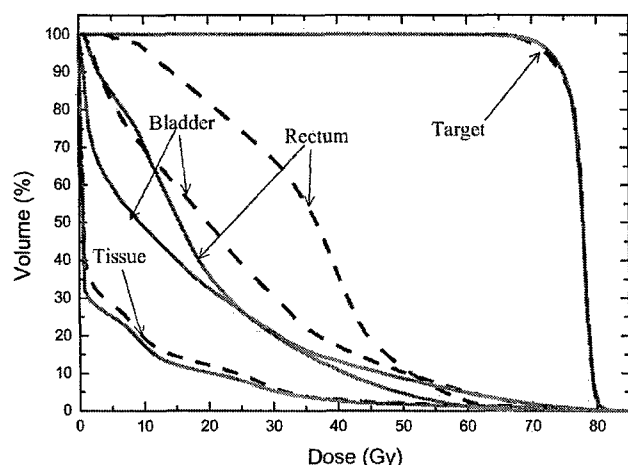


FIG. 4. Comparison of the dose volume histograms (DVH) for the prostate IMRT plans obtained using our newly proposed algorithm (solid curves) and the conventional algorithm (dashed curves).

verse slices and one sagittal slice for the two plans. The DVHs of the target and the sensitive structures for the two plans are plotted in Fig. 4, in which the dashed lines represent the results obtained using the conventional algorithm and the solid lines the newly proposed approach. Similar to the phantom case, when nonuniform importance factors are permissible, the conformality of the final dose distribution is greatly improved in comparison with the conventional IMRT planning with structurally uniform importance factors. In particular, the doses to the sensitive structures are dramatically improved. As can be seen from Fig. 3, much steeper dose gradients at the interface between the target and the sensitive structures are achieved. From the DVHs, it is observed that the rectum and bladder are better spared. For example, the fractional volume of the rectum that receives a dose above 40 Gy is dropped from 36.5% to 10.5%. Similarly, 10% reduction (from 30.5% to 20.5%) is noticed in the fractional bladder volume receiving a dose above 30 Gy. The integral dose to the normal tissue is, once again, reduced as a consequence of the use of nonuniform importance factors. It is important to emphasize that the improvement in sensitive structure sparing is achieved with essentially no change in the dose coverage of the tumor target. For the target, the fractional volume receiving dose below 70 Gy is decreased from about 2.1% to 1.2% and the fractional volume receiving dose above 80 Gy is decreased from 2.4% to 1.5%. The maximum doses for the two plans are almost the same, 81.2 Gy and 81.4 Gy, respectively. But the minimum dose is increased from 63.4 Gy for the conventional inverse planning to 65.5 Gy for the new technique.

Finally, it is interesting to point out that the EUD-based optimization²⁵⁻²⁷ is a special case of the empirical model proposed in this work. If we ignore the contribution from the dose discrepancy penalty part in Eq. (2) [i.e., change the power of the dose discrepancy in Eq. (2) from 2 to 0], the objective function (2) with r_i given by Eq. (3) becomes a function of EUD. Generally speaking, the formalism proposed here is a hybrid of dose-based and EUD-based func-

tions. Practically, the optimization problem here is formulated at a voxel level in dose domain while the EUD-based formalism works at a structural level in EUD domain. The voxel level information, such as nonuniform biology data in target and/or sensitive structures, can be more easily integrated into the objective function proposed here. In addition, prescription and evaluation of a treatment plan in dose domain other than in EUD domain makes the proposed method clinically more practical.

CONCLUSION

Inverse planning is essentially to balance the competing dosimetric or clinical requirements of the individual elements in a given system. The level of optimality of the final IMRT treatment plan depends not only on the performance of the optimization algorithm but also, more fundamentally, on the way that the tradeoff between different competing elements is defined. While a voxel in a patient is the smallest dosimetric element, the tradeoff in current inverse planning algorithm is being done implicitly on an organ or structure level. In this work, we have shown that this type of tradeoff scheme is deficient and seriously limits the solution space. Furthermore, we proposed a model for the tradeoff of different voxels through the modulation of voxel-dependent importance factors. A novel inverse planning algorithm with adaptively determined local importance factors was described. A comparison with the conventional inverse planning technique indicated that the new tradeoff strategy substantially improves the IMRT dose distributions and allows us to obtain solutions that would otherwise be unattainable.

ACKNOWLEDGMENTS

We wish to thank Dr. Zhenyu Shou for useful discussions. The support from the National Cancer Institute (1 R01 CA98523-01) and Department of Defense (DAMD17-03-1-0023) are gratefully acknowledged.

^{a)} Author to whom correspondence should be addressed; Electronic mail: lei@reyes.stanford.edu

¹ T. Bortfeld, "Optimized planning using physical objectives and constraints," *Semin Radiat. Oncol.* **9**, 20-34 (1999).

² S. V. Spirou and C. S. Chui, "A gradient inverse planning algorithm with dose-volume constraints," *Med. Phys.* **25**, 321-333 (1998).

³ Intensity Modulated Radiation Therapy Collaborative Working, "Intensity-modulated radiotherapy: current status and issues of interest," *Int. J. Radiat. Oncol., Biol., Phys.* **51**, 880-914 (2001).

⁴ R. Mohan, G. S. Mageras, B. Baldwin, L. J. Brewster, G. J. Kutcher, S. Leibel, C. M. Burman, C. C. Ling, and Z. Fuks, "Clinically relevant optimization of 3-D conformal treatments," *Med. Phys.* **19**, 933-944 (1992).

⁵ L. Xing, C. Cotrutz, S. Hunjan, A. L. Boyer, E. Adalsteinsson, and D. Spielman, "Inverse planning for functional image-guided intensity-modulated radiation therapy," *Phys. Med. Biol.* **47**, 3567-3578 (2002).

⁶ A. Brahme, "Optimized radiation therapy based on radiobiological objectives," *Semin Radiat. Oncol.* **9**, 35-47 (1999).

⁷ J. O. Deasy, "Multiple local minima in radiotherapy optimization problems with dose-volume constraints," *Med. Phys.* **24**, 1157-1161 (1997).

⁸ Y. Xiao, J. Galvin, M. Hossain, and R. Valicenti, "An optimized forward-planning technique for intensity modulated radiation therapy," *Med. Phys.* **27**, 2093-2099 (2000).

- ⁹S. Webb, "Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator," *Phys. Med. Biol.* **36**, 1201–1226 (1991).
- ¹⁰D. Michalski, Y. Xiao, Y. Censor, and J. Galvin, "The dose-volume constraint satisfaction problem for inverse treatment planning with field segments," *Phys. Med. Biol.* **49**, 601–616 (2004).
- ¹¹S. Gaede, E. Wong, and H. Rasmussen, "An algorithm for systematic selection of beam directions for IMRT," *Med. Phys.* **31**, 376–388 (2004).
- ¹²T. Bortfeld, C. Thieke, K. H. Kufer, and H. Trinkaus, "New approaches in intensity-modulated radiotherapy," ICRO Meeting, Monduzzi Editore, 2002, pp. 251–258.
- ¹³H. I. Amols and C. C. Ling, "EUD but not QUD," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 1–2 (2002).
- ¹⁴C. Yeboah, G. A. Sandison, and V. Moskvina, "Optimization of intensity-modulated very high energy (50–250 MeV) electron therapy," *Phys. Med. Biol.* **47**, 1285–1301 (2002).
- ¹⁵C. Cotrutz and L. Xing, "IMRT dose shaping with regionally variable penalty scheme," *Med. Phys.* **30**, 544–551 (2003).
- ¹⁶C. Cotrutz and L. Xing, "Using voxel-dependent importance factors for interactive DVH-based dose optimization," *Phys. Med. Biol.* **47**, 1659–1669 (2002).
- ¹⁷L. Xing, J. G. Li, S. Donaldson, Q. T. Le, and A. L. Boyer, "Optimization of importance factors in inverse planning," *Phys. Med. Biol.* **44**, 2525–2536 (1999).
- ¹⁸C. Wu, G. H. Olivera, R. Jeraj, H. Keller, and T. R. Mackie, "Treatment plan modification using voxel-based weighing factors/dose prescription," *Phys. Med. Biol.* **48**, 2479–2491 (2003).
- ¹⁹Y. Chen, D. Michalski, C. Houser, and J. M. Galvin, "A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy," *Phys. Med. Biol.* **47**, 1647–1658 (2002).
- ²⁰G. Starkschall, "A constrained least-squares optimization method for external beam radiation therapy treatment planning," *Med. Phys.* **11**, 659–665 (1984).
- ²¹T. Holmes and T. R. Mackie, "A comparison of three inverse treatment planning algorithms," *Phys. Med. Biol.* **39**, 91–106 (1994).
- ²²D. H. Hristov and B. G. Fallone, "An active set algorithm for treatment planning optimization," *Med. Phys.* **24**, 1455–1464 (1997).
- ²³L. Xing and G. T. Y. Chen, "Iterative algorithms for inverse treatment planning," *Phys. Med. Biol.* **41**, 2107–2123 (1996).
- ²⁴L. Xing, R. J. Hamilton, D. Spelbring, C. A. Pelizzari, G. T. Chen, and A. L. Boyer, "Fast iterative algorithms for three-dimensional inverse treatment planning," *Med. Phys.* **25**, 1845–1849 (1998).
- ²⁵A. Niemierko, "Reporting and analyzing dose distributions: a concept of equivalent uniform dose," *Med. Phys.* **24**, 103–110 (1997).
- ²⁶Q. Wu, R. Mohan, A. Niemierko, and R. Schmidt-Ullrich, "Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose," *Int. J. Radiat. Oncol., Biol., Phys.* **52**, 224–235 (2002).
- ²⁷C. Thieke, T. Bortfeld, A. Niemierko, and S. Nill, "From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning," *Med. Phys.* **30**, 2332–2339 (2003).

In Vivo Prostate Magnetic Resonance Spectroscopic Imaging Using Two-Dimensional J-Resolved PRESS at 3 T

Dong-hyun Kim,^{1*} Daniel Margolis,¹ Lei Xing,² Bruce Daniel,¹ and Daniel Spielman¹

In vivo magnetic resonance spectroscopic imaging of the prostate using single-voxel and multivoxel two-dimensional (2D) J-resolved sequences is investigated at a main magnetic field strength of 3 T. Citrate, an important metabolite often used to aid the detection of prostate cancer in magnetic resonance spectroscopic exams, can be reliably detected along with the other metabolites using this method. We show simulations and measurements of the citrate metabolite using 2D J-resolved spectroscopy to characterize the spectral pattern. Furthermore, using spiral readout gradients, the single-voxel 2D J-resolved method is extended to provide the spatial distribution information as well all within a reasonable scan time (17 min). Phantom and in vivo data are presented to illustrate the multivoxel 2D J-resolved spiral chemical shift imaging sequence. Magn Reson Med 53:1177–1182, 2005. © 2005 Wiley-Liss, Inc.

Key words: magnetic resonance spectroscopic imaging; prostate cancer; spiral readout gradients; 2D J-resolved spectroscopy; citrate; polyamine

In addition to the morphologic information provided by magnetic resonance imaging, the additional information gained using magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) increases the specificity of the examination for prostate cancer (PCa). In these examinations, the ratio of (choline + creatine) to citrate is often regarded as a marker for PCa (1). To date, MRSI protocols for PCa detection have been well established at a main magnetic field strength of 1.5 T (2).

The advent of higher field strength scanners provides the potential for improvement over 1.5-T systems due to the inherent increase in the signal-to-noise ratio (SNR). For PCa exams using MRS/MRSI methods, this advantage can be exploited in various forms, which include using higher spatial resolution acquisitions to increase the accuracy of localization of the cancerous tissues (3). Scan times can also be made shorter compared to 1.5 T for the same SNR, thereby reducing the overall MR examination time. The extension of 1.5-T MRS/MRSI protocols for usage in 3-T PCa can therefore have potential merits.

However, the process of advancing to higher field strength requires several considerations. For clinical prostate examinations using spectroscopic techniques, one of the issues that arise involves the detection of the citrate metabolite. Strong coupling of the AB system of citrate induces echo-time-dependent modulations of the signal response, which differs significantly with field strength (4,5). One method which exploits the echo time dependencies is the 2D J-resolved spectroscopic sequence (6–8). Acquisitions at incremental echo times can be gathered to obtain the coupling information of such metabolites. Information of uncoupled metabolites can also be gathered. Another advantage of using 2D J-resolved sequences in the case of prostate is the potential to separate the polyamine metabolite from the creatine and choline peaks (9,10). The three metabolites resonate at similar frequencies, which make them hard to differentiate using normal acquisitions. But, since the polyamines are also strongly coupled, the 2D J-resolved method can provide additional information, which can be used to distinguish between creatine and choline. In addition, 2D J-resolved spectroscopy has been used to reduce sideband artifacts for applications in the brain and regions outside of the brain such as the breast (11,12). Finally, the acquisitions from multiple echo times can also help determine the T_2 values of metabolites of interest in addition to water.

To take full advantage of this method, collecting the spatial distribution information of the metabolites will be preferred over single-voxel techniques. Although phase encoded MRSI methods have been used in obtaining the spatial information, this can be problematic when combining with the 2D J-resolved method since the minimum total scan time will be increased proportional to the number of echo time steps used. Therefore, a different approach that can reduce the minimum total scan time needs to be established. One of the spatial encoding methods that achieve such characteristic is the spiral readout MRSI (13). Using spiral MRSI, the spatial coverage can be controlled with high efficiency, thereby significantly decreasing the minimum total scan time compared to the phase encoded method. The additional time available can therefore be used to gather the 2D J-resolved spectroscopic data (14,15).

This work involves the study of 2D J-resolved single-voxel and multivoxel spectroscopic acquisition methods targeted for clinical application of PCa detection. We first explore the use of a 2D J-resolved single-voxel spectroscopic sequence to illustrate the detection of the citrate metabolite and to observe the characteristics of the J-coupled spectral pattern. Multivoxel 2D J-resolved acquisitions are performed using spiral-based MRSI.

¹Department of Radiology, Stanford University, Stanford, California, USA.

²Department of Radiation Oncology, Stanford University, Stanford, California, USA.

Grant sponsor: Lucas Foundation; Grant sponsor: NIH; Grant numbers: CA 48269, RR09784, and 1R01CA098523–01A1; Grant sponsor: Department of Defense; Grant number: DAMD17–03–1–0023.

*Correspondence to: Dong-hyun Kim, Radiological Science Laboratory at the Lucas MRS/I Center, Department of Radiology, Stanford University, 1201 Welch Road, Stanford, CA 94305–5488, USA. E-mail: dhkim@mrsc.ucsf.edu

Received 4 August 2004; revised 10 December 2004; accepted 15 December 2004.

DOI 10.1002/mrm.20452

Published online in Wiley InterScience (www.interscience.wiley.com).

© 2005 Wiley-Liss, Inc.

METHODS

Simulations and phantom measurements were conducted assuming a 2D J -resolved acquisition using a PRESS excitation scheme. For the simulations, we solved the full density matrix of strongly coupled two-spin systems with nonselective 180° pulses. The timing of the PRESS sequence was assumed to be $90^\circ - [t_{\text{int}}] - 180^\circ - [t_1/2] - 180^\circ - [t_1/2 - t_{\text{int}}] - t_2$ (acquire), where t_{int} was 10 ms. The J -coupling constant was assumed to be 15.4 Hz with a chemical shift value of 0.12 ppm (= 16.6 Hz at 3 T) (16). Also, the T_2 was assumed to be 200 ms with a line width of 10 Hz. For the actual single-voxel 2D J -resolved measurements, a phantom composed of citrate, creatine, and choline metabolites was used to emulate the existence of cancerous tissue. In both cases, the echo time spacing was adjusted to be 7.8 ms for a total of 64 steps from 35 to 534 ms in the F1 domain (Fourier domain corresponding to t_1 dimension). This resulted in a 2-Hz spectral resolution with a bandwidth of 128 Hz in the F1 domain. The spectral bandwidth in the F2 domain (Fourier domain corresponding to t_2 dimension) was 5000 Hz with 2048 data point acquisitions.

In vivo data were collected from a patient suspicious for recurrent PCa using the single-voxel 2D J -resolved technique. Single-voxel 2D J -resolved spectroscopic data using the PRESS sequence were acquired from two different regions near the peripheral zone of the prostate. All PRESS sequences were preceded by CHES water suppression and very selective saturation pulses for spatial saturation (17). The voxel size chosen was $1 \times 1.12 \times 1.08$ cm = 1.2 cc. Four acquisitions were averaged per incremental echo time for a total scan time of 8 min (TR = 2 s) for each voxel.

Multivoxel 2D J -resolved data using spiral MRSI were also acquired. A detailed description of the spiral MRSI

sequence can be found in Ref. 14. First, phantom data were collected to demonstrate the feasibility of obtaining spatial as well as the coupling information followed by in vivo data collection. Spiral readout gradients were applied to a PRESS excitation sequence. The spirals were designed using the formula given by Glover (18). A 32×32 spatial matrix covering a 24-cm field of view was used. Sixteen spatial interleaves of the spirals were used to acquire the required k -space. The number of spirals repeated during each acquisition was 256 lobes, which resulted in a 950-Hz spectral bandwidth in the F2 domain. Sixteen different echo times starting from 35 to 285 ms with 15.6-ms intervals were used to collect F1 domain information, which accounted for F1 spectral resolution of 4 Hz and a bandwidth of 64 Hz. Collected data were processed by first gridding in the k_x , k_y , and t_2 domains followed by a four-dimensional FT into the x , y , F1, and F2 domains. In summary, each of the 32×32 reconstructed voxels had 16 spectra, which covers the F1 dimension from $J(-32)$ Hz to $J(28)$ Hz. The nominal voxel size was 0.59 cc while the total scan time to acquire the necessary k -space and the t_1 space data was 17 min (TR = 2 s).

For the multivoxel 2D J -resolved experiment, phantom data were acquired using the head coil while in vivo data were collected with the body coil for excitation followed by a rigid endorectal coil for signal reception. A phantom comprised of creatine, choline, and citrate solution was built for this study, which was surrounded by lipids to confirm the lipid suppression capabilities. For the in vivo exams, a high-resolution (512×512) T_2 -weighted anatomic image was obtained and a region of interest covering most of the prostate was selected for the PRESS sequence. To date, seven patients who were suspicious of prostate cancer were referred to by a pathologist and imaged using

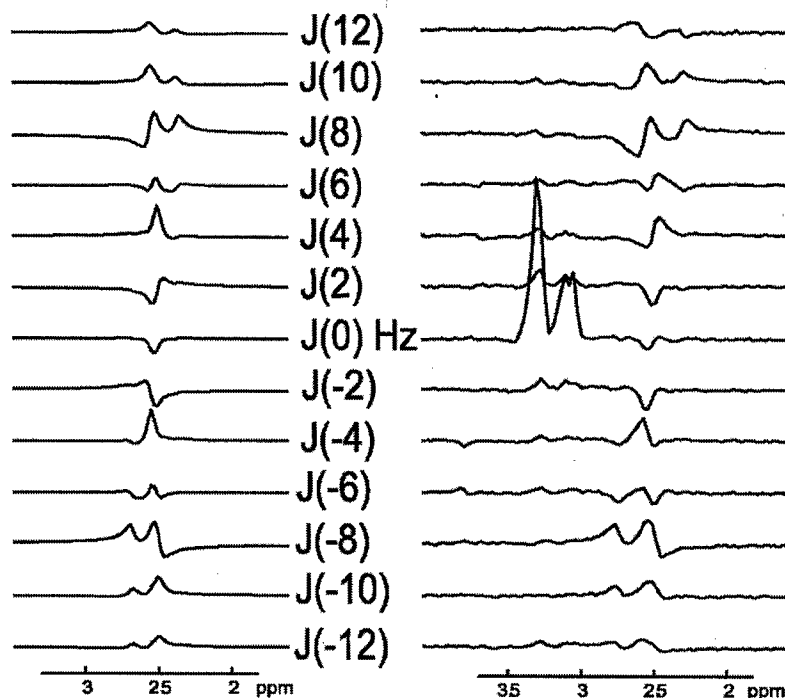


FIG. 1. 2D J -resolved spectra obtained from simulations (left) and phantom measurements (right) at 3 T. The echo time interval was 7.8 ms starting from 35 ms for 64 steps. Choline and creatine metabolites were added for the phantom measurement. Due to the modulations as a function of echo time, the 2D J -resolved spectra show a strong signal from the citrate metabolite at reconstructed lines beyond the $J(0)$ Hz line for both simulated and measured data with similar spectral patterns. In this respect, the detection of the citrate resonance can be made outside of the $J(0)$ Hz line using the 2D J -resolved acquisition.

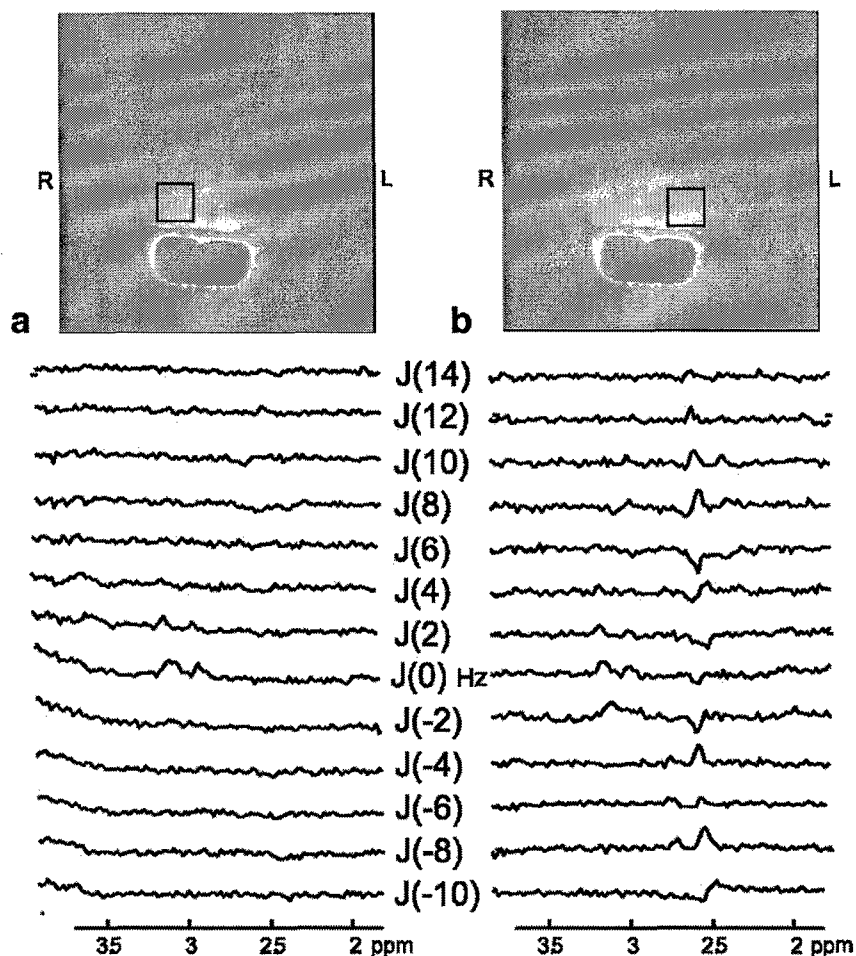


FIG. 2. Single-voxel 2D *J*-resolved spectroscopy results obtained in vivo from a subject suspicious of recurrent PCa. Two voxels were selected for the examination as shown in the T_2 -weighted images. The reconstructed spectra of several F1 lines are shown. In (a), even though the presence of creatine and choline metabolites is evident, there is no visible citrate. As for the region shown in (b), the citrate is visible (2.6 ppm region from $J(-10)$ to $J(12)$ Hz line) in the spectra while other metabolites are also present. This shows that the 2D *J*-resolved spectroscopy can be useful for in vivo detection of citrate.

a 3 T GE Signa scanner (GE Health Care, Waukesha, WI). All in vivo studies were conducted under IRB guidelines and with informed consent.

RESULTS

Figure 1 shows the simulated 2D *J*-resolved citrate spectra (left) along with the reconstructed 2D *J*-resolved data acquired with a phantom (right) using the single-voxel 2D *J*-resolved acquisition. Spectra corresponding to the F1 domain in the range of $J(-12)$ to $J(12)$ Hz were extracted where most of the energy is concentrated. In both cases, due to the modulations occurring as a function of echo time, resonances are clearly seen beyond the $J(0)$ line for the citrate metabolite. Individual spectra from each F1 line reveal the similarity of the patterns between the simulated and measured results of the citrate. The $J(0)$ line, also referred to as the TE-averaged line, has a slight negative peak at the citrate position, which is due to the strong negative peaks at echo times ranging from 60 to 120 ms.

Two single-voxel 2D *J*-resolved spectra from an in vivo subject are presented in Fig. 2. The patient had a prior history of prostatic adenocarcinoma, which was treated by external beam radiation. The two regions that were selected are shown in the anatomic T_2 -weighted images along with the resulting *J*-resolved spectra. The spectra

obtained from the right side of the subject (Fig. 2a) displays negligible citrate metabolite intensity compared to the creatine and choline resonances located near the 3.0-ppm region. In comparison, the spectra from the left side of the subject (Fig. 2b) reveal the presence of citrate as seen from the modulations occurring in the reconstructed F1 lines along with the creatine and choline metabolites. These two comparisons show that with the 2D *J*-resolved acquisition method, the strongly coupled citrate metabolite can be resolved while the presence of other metabolites can be established. Even though the number of radio-frequency (RF) phase cycling steps has been reduced to 4 in this case, strong residual signal from outside of the PRESS box is not observed.

In Fig. 3, results obtained from the multivoxel 2D *J*-resolved sequence via spiral MRSI are shown. In Fig. 3a, an image of the phantom that was used for the experiment is given. In Fig. 3b, the metabolite spectra corresponding to the voxel selected in Fig. 3a are given. We extracted the TE-averaged line from each reconstructed voxel and manually phased them. The TE-averaged spectra show the well-resolved spatial distribution of the metabolites with a slight negative peak of the citrate as in the case of the single-voxel experiment. Spatial saturation pulses eliminated most of the lipids, as can be seen from the absence of any sidebands arising from the lipids. In Fig. 3c, lines from

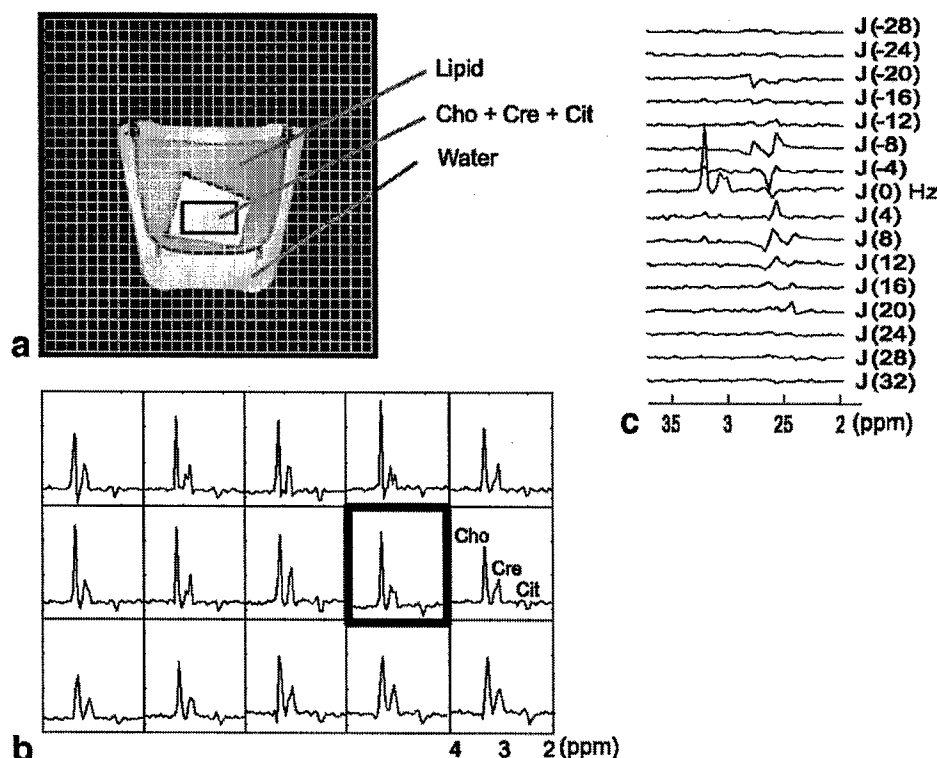


FIG. 3. Phantom results using spiral based multivoxel 2D J -resolved MRSI. A metabolite phantom surrounded by lipid and water was used (a). In (b), the metabolite spectra corresponding to $J(0)$ Hz are displayed from the selected voxels shown in (a). The TE-averaged spectra show the well-resolved spatial distribution of the metabolites. In (c), all the reconstructed F1 lines for the voxel highlighted in (b) are shown. The F1 domain lines clearly depict the presence of citrate near the 2.5-ppm region.

$J(-28)$ to $J(32)$ Hz that were reconstructed for the voxel highlighted in Fig. 3b are shown. The J -resolved F1 domain lines clearly depict the presence of citrate located near the 2.5-ppm region.

Figure 4 shows spiral readout 2D J -resolved MRSI results obtained from an in vivo subject who had been reported as having a Gleason score of 3 + 3. The T_2 -weighted image is shown in Fig. 4a with a grid representing the displayed voxels shown in Fig. 4b and c. Reconstructed spectra corresponding to the TE-averaged lines are given in Fig. 4b. As with the case of the phantom experiment, the TE-averaged line largely represents spectra from metabolites that are uncoupled. This is illustrated by the existence of choline and creatine metabolites that can be seen near the middle region of the displayed voxels. In Fig. 4c, the spectra corresponding to $J(8)$ Hz are displayed where the citrate metabolite can be resolved. This is illustrated near the upper left region of the prostate where several voxels show a peak near the 2.5 ppm region, which corresponds to the citrate metabolite. For several voxels, however, lipid contamination can be visible near the 2.5-ppm region, which compromised the detection of citrate.

Figure 5 shows spiral readout 2D J -resolved MRSI results obtained from another in vivo subject who had been reported as having adenocarcinoma of the prostate with a Gleason score of 3 + 4. The TE-average lines in Fig. 5a show voxels of signal contributing from creatine and choline as well as polyamines residing in $J(0)$ Hz. In Fig. 5c, the spectra corresponding to $J(8)$ Hz line are displayed. The spectra show clear visualization of the citrate metabolite. In addition, polyamines are seen to be resolved as well. Of the seven patients examined, five patients had observable signal from any of the metabolites of interest.

DISCUSSION

We have shown the application of a 2D J -resolved PRESS sequence, which can aid the detection of PCa at the field strength of 3 T. In the first part of the paper, simulations and phantom measurements showing the characteristics of the citrate metabolite resonance for 2D J -resolved single-voxel acquisitions were illustrated, which showed the strong dependency with echo time due to its strong coupling. In the second part spiral readout gradients were applied to the 2D J -resolved PRESS sequence to obtain additional spatial distribution information. The efficiency of the spiral k -space trajectory makes it possible to cover the whole k -space within a reasonable scan time.

For a truly feasible clinical protocol to be implemented, several prerequisites need to be established, however. It is important for a good spatial suppression pulse to be used so that there are no aliasing or ringing artifacts. This problem has been demonstrated from the in vivo exam where lipid sidebands interfered with the detection of metabolites. This result is in comparison with the phantom results where good lipid suppression was accomplished. For the phantom experiment, a head coil with a relatively homogeneous RF profile and low power was used, whereas this was not the case for the in vivo exam, which used the body coil for excitation. Another important prerequisite is the main field homogeneity. The presence of air inside the endorectal coil or near the prostate region can degrade the homogeneity, leading to line broadening and potential overlap of the choline and creatine metabolites as seen in the in vivo example.

In this study, we addressed the issue of strong coupling citrate peak using a 2D J -resolved spectroscopic acquisi-

FIG. 4. Results obtained from a patient diagnosed with prostate cancer with a Gleason score 3 + 3 using spiral-based 2D J-resolved MRSI. (a) T_2 -weighted image with a grid representing the displayed voxels. (b) Reconstructed spectra corresponding to the $J(0)$ Hz lines. The existence of choline and creatine metabolites can be seen near the middle region of the displayed voxels. (c) Spectra corresponding to $J(8)$ Hz line from F1 domain. Several voxels show a peak near the 2.5-ppm region at the upper left region, which corresponds to citrate. Lipid contamination can be visible near the 2.5-ppm region for several voxels.

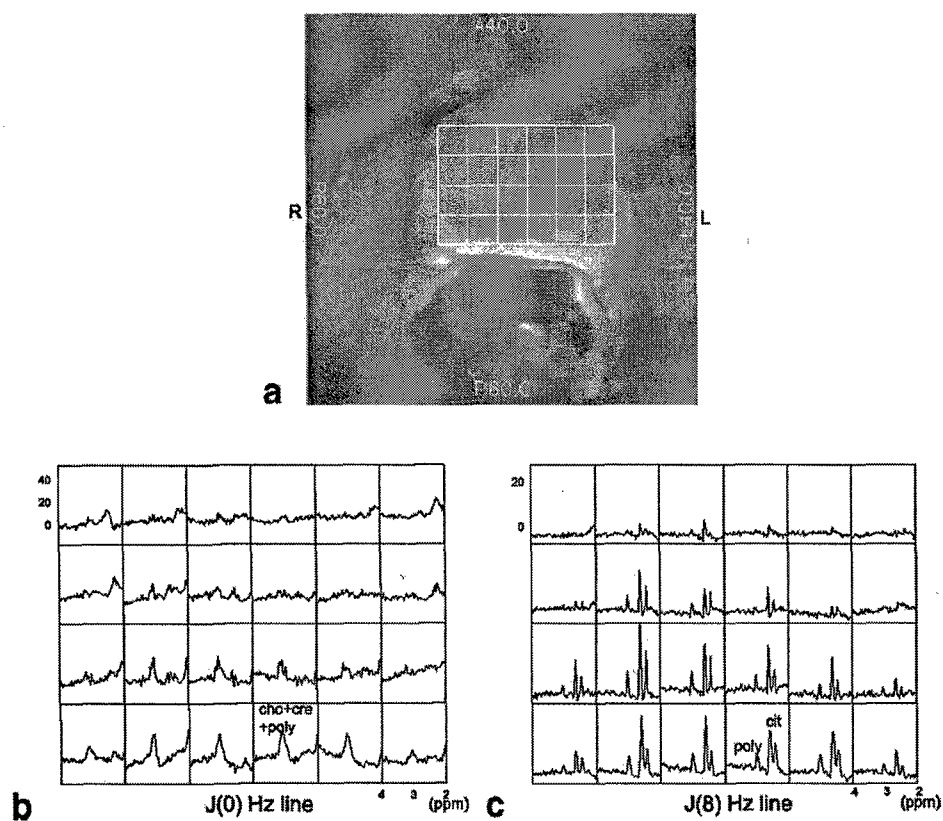
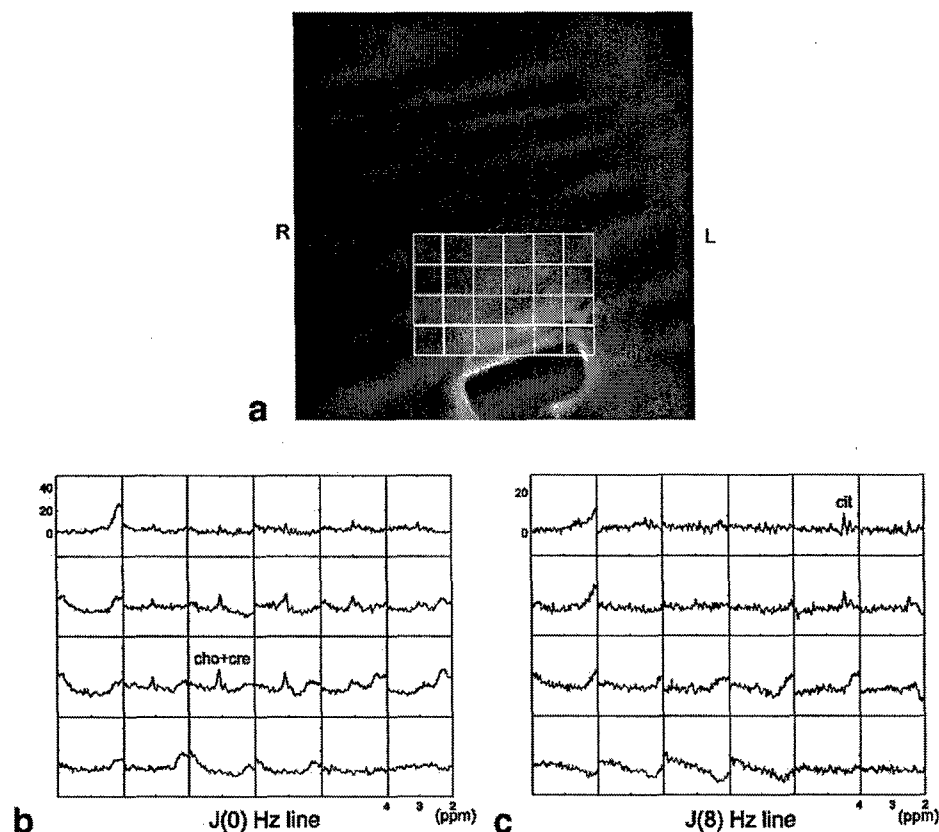


FIG. 5. Results obtained from a patient diagnosed with adenocarcinoma of the prostate with a Gleason score 3 + 4 using spiral-based 2D J-resolved MRSI. (a) T_2 -weighted image with a grid representing the displayed voxels. (b) Reconstructed spectra corresponding to the $J(0)$ Hz lines. The coexistence of choline, creatine, and polyamine metabolites can be seen. (c) Spectra corresponding to the $J(8)$ Hz line. Several voxels show a clear peak near the 2.5-ppm region corresponding to citrate. Polyamines can also be observed for several voxels around the 3.1-ppm region.

tion sequence. This can be dealt with in a different way as recently shown using a *J*-refocused sequence (19). One of the advantages of using the 2D *J*-resolved method includes the potential to detect changes in the citrate coupling constant, which can be used as another marker for PCa. This coupling constant is believed to be related to the zinc concentration, which is directly related to presence of PCa (20). In addition, any information obtained from the second spectral dimension, for example, from the polyamines as seen, can add to the physiologic information of the prostate tissue (9). On the negative side, a *J*-refocused scheme would require many fewer acquisitions and can achieve better SNR due to a shortened TE.

CONCLUSION

Single-voxel and multivoxel 2D *J*-resolved spectroscopy methods have been demonstrated for in vivo prostate at field strength of 3 T. Using the 2D *J*-resolved method, strong coupling of citrate can be well resolved. For multivoxel 2D *J*-resolved spectroscopic imaging, spiral-based readout sequences are used, which enable data acquisition within a reasonable scan time.

REFERENCES

1. Scheidler J, Hricak H, Vigneron DB, Yu KK, Sokolov DL, Huang RL, Zaloudek CJ, Nelson SJ, Carroll PR, Kurhanewicz J. 3D ¹H-MR spectroscopic imaging in localizing prostate cancer: clinico-pathologic study. *Radiology* 1999;213:473-480.
2. Kurhanewicz J, Vigneron DB, Hricak H, Narayan P, Carroll P, Nelson SJ. Three-dimensional ¹H MR spectroscopic imaging of the in situ human prostate with high (0.24-0.7 cm³) spatial resolution. *Radiology* 1996;198:795-805.
3. Vigneron DB, Chen A, Cunningham C, Xu D, Hurd R, Sailasuta N, Pauly J, Nelson S, Kurhanewicz J. High resolution 3D MR spectroscopic imaging and *J*-resolved MRS of the prostate at 3 Tesla. In: Proceedings of the 12th Annual Meeting of ISMRM, Kyoto, Japan, 2004. p 386.
4. Wilman AH, Allen PS. The response of the strongly coupled AB system of citrate to typical ¹H MRS localization sequences. *J Magn Reson B* 1995;107:25-33.
5. Mulkern RV, Bowers JL, Peled S, Kraft RA, Williamson DS. Citrate signal enhancement with a homonuclear *J*-refocusing modification to double-echo PRESS sequences. *Magn Reson Med* 1996;36:775-780.
6. Aue WP, Karhan J, Ernst RR. Homonuclear broadband decoupling and two-dimensional *J*-resolved NMR spectroscopy. *J Chem Phys* 1976;64:4226.
7. Thomas MA, Ryner L, Mehta M, Turski P, Sorenson JA. Localized 2D *J*-resolved ¹H MR spectroscopy of human brain tumors in vivo. *J Magn Reson Imaging* 1996;6:453-459.
8. Dreher W, Leibfritz D. On the use of two-dimensional-*J* NMR measurements for in vivo proton MRS: Measurements of homonuclear decoupled spectra without the need for short echo time. *Magn Reson Med* 1995;34:331-337.
9. Swanson MG, Vigneron DB, Tran TC, Sailasuta N, Hurd RE, Kurhanewicz J. Single-voxel oversampled *J*-resolved spectroscopy of in vivo human prostate tissue. *Magn Reson Med* 2001;45:973-980.
10. Yue K, Marumoto A, Binesh N, Thomas MA. 2D JPRESS of human prostates using an endorectal receiver coil. *Magn Reson Med* 2002;47:1059-1064.
11. Hurd RE, Gurr D, Sailasuta N. Proton spectroscopy without water suppression: the oversampled *J*-resolved experiment. *Magn Reson Med* 1998;40:343-347.
12. Bolan PJ, DelaBarre L, Baker EH, Merkle H, Everson LI, Yee D, Garwood M. Eliminating spurious lipid sidebands in ¹H MRS of breast lesions. *Magn Reson Med* 2002;48:215-222.
13. Adalsteinsson E, Irarrazabal P, Topp S, Meyer C, Macovski A, Spielman DM. Volumetric spectroscopic imaging with spiral-based k-space trajectories. *Magn Reson Med* 1998;39:889-898.
14. Adalsteinsson E, Spielman DM. Spatially resolved two-dimensional spectroscopy. *Magn Reson Med* 1999;41:8-12.
15. Hiba B, Serduc R, Provent P, Farion R, Remy C, Ziegler A. 2D *J*-resolved spiral spectroscopic imaging at 7 T: application to mobile lipid mapping in a rat glioma. *Magn Reson Med* 2004;52:658-662.
16. Schick F, Bongers H, Kurz S, Jung W, Pfeffer M, Lutz O. Localized proton MR spectroscopy of citrate in vitro and of the human prostate in vivo at 1.5T. *Magn Reson Med* 1993;29:38-43.
17. Tran TC, Vigneron D, Sailasuta N, Tropp J, LeRoux P, Kurhanewicz J, Nelson SJ, Hurd R. Very selective suppression pulses for clinical MRSI studies of brain and prostate cancer. *Magn Reson Med* 2000;43:23-33.
18. Glover GH. Simple analytic spiral k-space algorithm. *Magn Reson Med* 1999;42:412-415.
19. Cunningham CH, Marjanska M, Chen AP, Xu D, Pauly JM, Sailasuta N, Hurd RE, Kurhanewicz J, Garwood M, Vigneron DB. Sequence design incorporating the LASER technique for prostate MRSI at high field. In: Proceedings of the 12th Annual Meeting of ISMRM, Toronto, Canada, 2003. p 682.
20. Van der Graaf M, Heerschap A. Effect of cation binding on the protonchemical shifts and the spin-spin coupling constant of citrate. *J Magn Reson B* 1996;112:58-62.

Feasibility study of beam orientation class-solutions for prostate IMRT

Eduard Schreibmann and Lei Xing^{a)}

*Department of Radiation Oncology, Stanford University School of Medicine,
Stanford, California 94305-5847*

(Received 2 April 2004; revised 22 July 2004; accepted for publication 4 August 2004;
published 30 September 2004)

IMRT is being increasingly used for treatment of prostate cancer. In practice, however, the beam orientations used for the treatments are still selected empirically, without any guideline. The purpose of this work was to investigate interpatient variation of the optimal beam configuration and to facilitate intensity modulated radiation therapy (IMRT) prostate treatment planning by proposing a set of beam orientation class-solutions for a range of numbers of incident beams. We used fifteen prostate cases to generate the beam orientation class-solutions. For each patient and a given number of incident beams, a multiobjective optimization engine was employed to provide optimal beam directions. For the fifteen cases considered, the gantry angle of any of the optimized plans were all distributed within a certain range. The angular distributions of the optimal beams were analyzed and the most selected directions are identified as optimal directions. The optimal directions for all patients are averaged to obtain the class-solution. The class-solution gantry angles for prostate IMRT were found to be: three beams (0° , 120° , 240°), five beams (35° , 110° , 180° , 250° , 325°), six beams (0° , 60° , 120° , 180° , 240° , 300°), seven beams (25° , 75° , 130° , 180° , 230° , 285° , 335°), eight beams (20° , 70° , 110° , 150° , 200° , 250° , 290° , 340°), and nine beams (20° , 60° , 100° , 140° , 180° , 220° , 260° , 300° , 340°). The level of validity of the class-solutions was tested using an additional clinical prostate case by comparing with the individually optimized beam configurations. The difference between the plans obtained with class-solutions and patient-specific optimizations was found to be clinically insignificant. © 2004 American Association of Physicists in Medicine. [DOI: 10.1118/1.1797571]

I. INTRODUCTION

Beam configuration may have significant influence on the quality of an IMRT treatment.¹ Clinically, however, the gantry angles are still selected empirically and multiple adjustments may often be required. This has become one of the bottleneck problems that hinder the optimal and efficient use of IMRT. A "class-solution" method for beam placement has been considered.² The idea behind this approach is to construct a representative beam configuration based on previous experience, or more adequately, on accurate beam orientation optimization, for a given disease site and then use this class-solution for subsequent patient treatment planning. This paper systematically investigates the issue and proposes a set of beam orientation class-solutions for IMRT prostate irradiation. The importance of class-solutions has been apparently recognized also by treatment planning system (TPS) vendors. For example the Corvus system has the functionality to define and use class solutions, the challenge of identifying optimal directions is however left to on the user.

In the past few years, several groups have investigated the problem of beam orientations in three-dimensional conformal radiation therapy³⁻¹³ and in IMRT.^{9,14-22} The main focus of these studies was on the modeling, formulation, and optimization of beam configurations. To obtain a set of good gantry angles for an IMRT treatment, in principle, all one needs to do is to add the gantry angle variables into an objective function and then to optimize the objective function with respect to the gantry angles and the beamlet weights. A main difficulty of this type of brute-force optimization lies in

its excessive computational time caused by the greatly enlarged search space. The development of beam orientation class-solution for some simple disease sites like prostate has the potential to significantly reduce the effort of beam placement in IMRT planning, which is thus of practical importance.

Validity of the class-solution approach has been demonstrated by Nutting *et al.*²³ for parotid gland tumors cases. To derive a class-solution, locations of optimal beams should not vary significantly among different patients. If such a condition is met, the individual patient optimization process can be skipped in clinical practice with a minimal compromise. Trials to manually derive a suitable set of constraints and beam orientations for prostate cases have also been conducted.^{24,25}

To derive a population-based beam orientation class-solution, the optimal solutions need to be obtained for at least a representative group of patients. For this purpose, in the next section we will present an effective beam orientation algorithm for the optimization of IMRT beam orientations. It is intuitively conceivable that the optimal beam configuration varies from patient to patient because of their geometric differences. Generally speaking, a class-solution exists if the interpatient variation in the beam orientations is small. As will be seen in Sec. III this is approximately the case for the representative group of fifteen prostate cases selected randomly from the patient pool treated at Stanford. We have investigated the behavior of the system for a variety of numbers of incident beams and provided the class-solutions for

the cases of three, and five to nine incident beams to meet different clinical needs. Our study has indicated that the beam orientation class-solutions provided here represent good compromise between practicality and optimality and may facilitate the IMRT treatment planning by eliminating or simplifying the trial-and-error process of the beam orientation selection.

II. MATERIALS AND METHODS

A. Multiobjective optimization

A multiobjective (MO) optimization with a mix of genetic and iterative algorithms was used to optimize the system. The genetic algorithm optimizes the beam orientations and the gradient-based method the intensity profiles of the beams. The MO is useful in radiotherapy dose optimization because of the existence of multiple conflicting objectives in the system. In the previous studies, the problem has been transformed into a single objective problem by introducing a set of importance factors to parameterize our tradeoff strategy.²⁶ The MO tries to obtain all acceptable solutions and provides more options on the tradeoff between the different objectives. Here, by an acceptable solution we mean a plan with a good compromise of all the objectives involved in the problem. Mathematically, the MO optimization or vector optimization²⁷ is to determine a set of decision variables that optimizes a vector function whose elements represent M objective functions without violating the system constraints.

We denote the decision variables in an optimization problem by $\mathbf{x} = \{x_j, j = 1, 2, \dots, N\}$. To access the goodness of a solution it is needed to establish some criteria of evaluation. These criteria are expressed as computable functions $f_1(\mathbf{x}), \dots, f_M(\mathbf{x})$ of the decision variables, which are called *objective functions*. These form a vector function \mathbf{f} . The vector function $\mathbf{f}(\mathbf{x})$ maps the set \mathbf{X} to the set \mathbf{F} that represents all possible values of the objective functions. In general, some of these objectives will be in conflict with others and do not have a situation in which all the $f_i(\mathbf{x})$ reach their optimums at a common point \mathbf{x} . The multiobjective optimization problem is to find the vector $\mathbf{x} = (x_1, x_2, \dots, x_N)$, i.e., the solution which optimize the vector function \mathbf{f} .

To proceed, it is necessary to establish some criteria to define the optimality. The most commonly adopted definition of "optimality" is the Pareto optimum based on the dominance relation. A solution \mathbf{x}_1 dominates a solution \mathbf{x}_2 if the two following conditions are satisfied:

- \mathbf{x}_1 is no worse than \mathbf{x}_2 in all objectives, i.e., $f_j(\mathbf{x}_1) \leq f_j(\mathbf{x}_2) \forall j = 1, \dots, M$,
- \mathbf{x}_1 is strictly better than \mathbf{x}_2 in at least one objective, i.e., $f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2)$, for at least one $j \in \{1, 2, \dots, M\}$.

We assume, without loss of generality that the problem that we are dealing with here is a minimization problem. Among a set of solutions P , the nondominated set of solutions P' are those that are not dominated by any other member of the set P . When the set P is the entire feasible search space, then the set P' is called the *global Pareto optimal set*.

If for every member \mathbf{x} of a set P there exists no solution in the neighborhood of \mathbf{x} , then the solutions of P form a *local Pareto optimal set*. The image of the Pareto optimal set is called the *Pareto front*.

B. Optimization method

The variables to be optimized include the beam directions and the bixel intensities. We quantify the variables in a two-component chromosome. The first part \mathbf{W} contains the bixel weights of all beams with a double precision floating-point representation. The second part \mathbf{B} is a boolean string, which represents the possible beam directions. The boolean value records if the beam is active or not.

The system was optimized using two algorithms: a genetic algorithm as the main optimization framework and a gradient-based algorithm applied on a fraction of the newly generated solutions. The algorithm was described in detail in a previous publication,²⁸ and a similar optimization approach was recently reported by another group.²⁹ Without the gradient-based algorithm, the genetic algorithm will converge very slowly because the cost function is too sensitive to the changes in bixel weights generated by the genetic algorithm. The gradient-based algorithm, used for a few iterations on the new solutions, lowers the sensitivity by adjusting the bixel weights. The MO multiobjective nondominated sorting genetic algorithm with constrained elitism *NSGA-II/C*,³⁰ which is one of the most effective evolutionary optimization algorithms, was used as the genetic calculation engine. The selected gradient-based algorithm was the low memory BFGS algorithm (L-BFGS).³¹ The crossover and mutation probability were chosen as 0.9 and 0.01, respectively for this study. The algorithm used the arithmetic crossover and the nonuniform mutation. Due to the large number of variables, we used a population size of 100 to maintain the genetic algorithm's diversity. These optimal parameters have been estimated by comparing with results obtained using a deterministic gradient-based algorithm, which indicated that the Pareto front does not change significantly after 100 generations.³²

In this study, the beams of coplanar setup were divided in 360 beams with 1° separation between adjacent beams. Gantry motion is specified according to IEC convention. Each beam was divided in a grid of 40×40 bixels and each bixel had a dimension of $1 \times 1 \text{ cm}^2$, according to the collimator leaf size used in our institution and to the bixel size used in commercial TPS. Only those bixels that pass through the planned target volume (PTV) were activated and considered in the optimization process. The number of bixels to be optimized depended on the patient's geometry and number of beams, ranging from 4050 to 9584 for the prostate cases analyzed in this paper. All calculations were performed in three-dimensional geometry with 15-MV photons. A Clarkson dose computation model was used. The dose value at each sampling point was obtained from the interpolation of tissue phantom ratio values and off-center ratio values. The objective function was computed over quasirandom generated sampling points for each structure,³³ at a density of

TABLE I. Dose limits and volume limits used as constraints for OARs. The prescribed dose was 72 Gy to the PTV and 0 Gy to the OARs. For each OAR, two constraints are imposed: (a) the maximum dose and (b) a specified volume should not receive a dose greater than the limit dose.

Organ	D_{max} (Gy)	D_{Critic} (Gy)	V_{Critic} (% of organ volume)
NT	75	40	20
Bladder	65	50	50
Femurs	68	40	20
Rectum	72	53	25

30 points/cm³. A minimum and maximum limit of 1000 and 30 000 points were used to avoid undersampling/oversampling of small/large structures.

C. Objective functions

The objective functions used in this study for the PTV and for organs at risk (OAR) are defined as

$$f_{PTV} = \frac{1}{N_{PTV}} \sum_{j=1}^{N_{PTV}} (d_j^{PTV} - D_{ref})^2, \quad f_{OAR} = \frac{1}{N_{OAR}} \sum_{j=1}^{N_{OAR}} (d_j^{OAR})^2,$$

where D_{ref} is the prescription dose to the PTV, d_j^{PTV} and d_j^{OAR} are the calculated doses at the j th sampling point for the PTV and OAR, respectively, and N_{PTV} and N_{OAR} are the corresponding number of sampling points.

D. Dosimetric constraints

We imposed maximum dose and dose-volume constraints on OARs. The maximum dose constraint was defined as: "maximum dose should be below D_{Limit} ." The dose-volumetric constraint was formulated as "no more than V_{Critic} should be receive doses higher than D_{Critic} ." The values of D_{Limit} , V_{Critic} , and D_{Critic} were assigned according to the literature^{34,35} and are summarized in Table I. For the PTV, imposed constraints were: (a) the maximum dose should be less than 78 Gy; and (b) at least 95% of the PTV should receive 72 Gy.

E. Selection of solutions from the Pareto set

A set of solutions is generated at the end of a multiobjective optimization, each corresponding to an IMRT plan that is relevant to the clinical treatment. Specifically, the treatment planner is provided with a table of values for all the members of the Pareto set of the objectives, dose volume histograms (DVHs) for all OARs and PTV, and dose statistics. For a given patient and pre-selected number of beams, an appropriate solution is generally chosen from the Pareto set based on clinical considerations.^{17,36} The optimization objectives were to conform the PTV, while minimizing the dose to bladder, rectum, and femoral heads.

F. Beam orientation class-solutions

Fifteen prostate cases were selected randomly from the patient pool at Stanford University Hospital. The patients were scanned in supine position, using a 3 or 4 mm CT slice thickness. The contours were delineated on a Philips virtual simulation system (Philips Medical System, Cleveland, OH) and imported in our in-house treatment planning system using the DICOM transfer protocol. The beam orientation optimization was then performed for each of the patients for a range of number of incident beams ($N=3, 5, 6, 7, 8$, and 9) using the algorithm described above. Only coplanar beams were considered in this work. The angular distribution of the incident beams for a given number of incident beams for the fifteen patients was analyzed to identify optimal directions. For each number of beams, we identified the directions for the class-solution by averaging the optimal directions of the fifteen patients.

G. Test of the degree of validity of class-solution

In order to test the validity of the beam orientation class-solution approach, we randomly selected an additional prostate patient from the patient pool at Stanford University Hospital. Two IMRT plans were generated for the patient: one with class-solution selected beams and one with beams chosen by the patient-specific beam orientation optimization. The difference between the two plans is analyzed according to the commonly used plan evaluation indices, such as DVH, maximum/minimum/average target/sensitive structure doses.

III. RESULTS

A. Pareto front plans for a given patient

For a given patient, a set of plans is provided by the multiobjective optimization described in the last section, each represents a solution in the Pareto front. The DVH of a given structure in multiobjective optimization constitutes a spectrum of curves instead of a single curve. In Fig. 1 we show the DVH spectra of prostate, rectum, and bladder for a representative patient for all the nondominated five-field IMRT plans in the Pareto front.

Each nondominated plan has different beam parameters (bixel maps and/or gantry angles). The angular distributions of the Pareto front plans for the above patient are shown in Fig. 2 for different numbers of incident beams. If a beam direction is more feasible it shows up in the final plans with high frequency. For three beams the directions that appeared with high frequencies in the Pareto front plans are at 120°, 240°, and 340°. It is interesting to note that the beam configuration is close to that found by inspecting the most-selected plan. Small peaks appear at 60° and 290° because they are opposite to the most favorable beam directions, i.e., 240° and 120°.

As will be seen in the next section, for three to seven beams, the angles of the maximums of the angular distributions are close to the directions found using the most-selected plans. As the number of beams increased, the importance of optimizing beam orientation decreases and a large

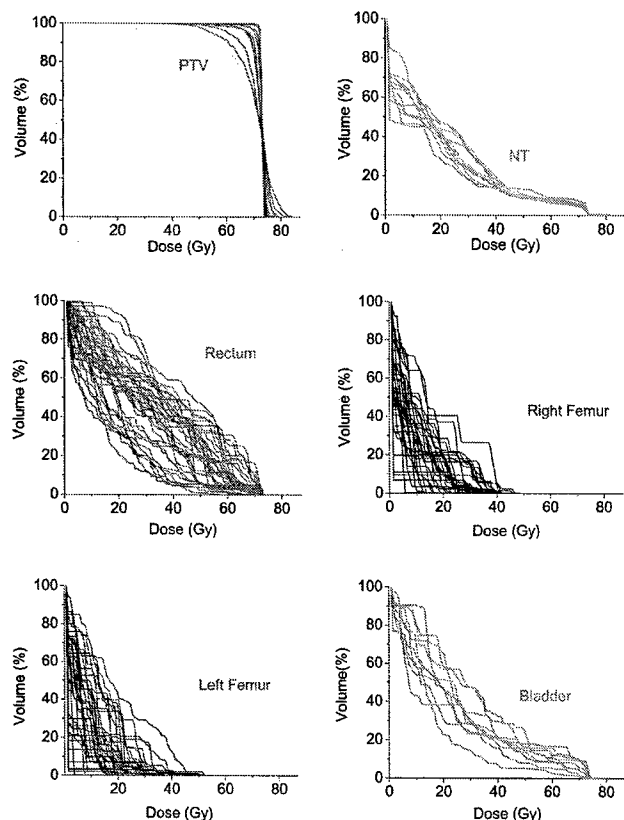


FIG. 1. DVH spectra of OARs and PTV for a set of clinically relevant nondominated plans in the Pareto front. The result of a multiobjective optimization is a set of plans (Pareto front) describing the "optimal of the optimized plans," i.e., no objective can be improved without degrading another one.

set of beam configurations with more "diffusive" angular distributions become capable of producing valid plans with adequate PTV coverage and OAR sparing. In this situation, the feature of distinct peaks in the angular distribution diminishes.

B. Class-solutions

For a given patient, the beam configuration occurring most frequently in the Pareto front is identified as the optimal one. In Fig. 3 we show the optimized five-beam configurations for each of the fifteen patients. It is intriguing to note that all cases have a similar beam setup, with almost equispaced beams starting at the beam position of 20° – 45° . To better visualize the result, in Fig. 4(b) we show the gantry angles for the fifteen patients (Fig. 4) on a single plot. The gantry angles of the class-solution for IMRT treatment with five incident beams seems to be 35° , 110° , 180° , 250° , and 325° , and are indicated in the top lines in Fig. 4(b). For an individual patient, optimal directions are identified at high frequency locations in the angular distributions shown in Figs. 3 and 4. Class solutions directions are found by averaging optimal directions of all fifteen patients. For practice purposes, one may want to "round" the beam angles to a set of more convenient implementable beams. With the consideration of angular distribution of each beam shown in Fig. 4(b), a set of equispaced beams starting from 20° – 45° should also provide a good compromised solution because of only a small adjustment is done with the range of the angular distributions of the beams.

The solutions derived above seem to be physically sensible. Setting the angles of beams 1, 2, and 4 to 35° , 110° , and 250° , respectively balances the dose to the femoral heads. The locations of the third (180°) and fifth (325°) beams are chosen to balance the dose requirements of the PTV and rectum. Depending on the geometry of the case, the optimizer makes individual adjustments of the beam angles, but the selected directions are within certain angular ranges, which are 40° , 40° , 40° , 34° , and 52° .

The same calculation described above is repeated for three, and six through nine incident beams, and the distributions of the optimal configurations for the fifteen cases are shown in Fig. 4. The class-solution for each of the situations

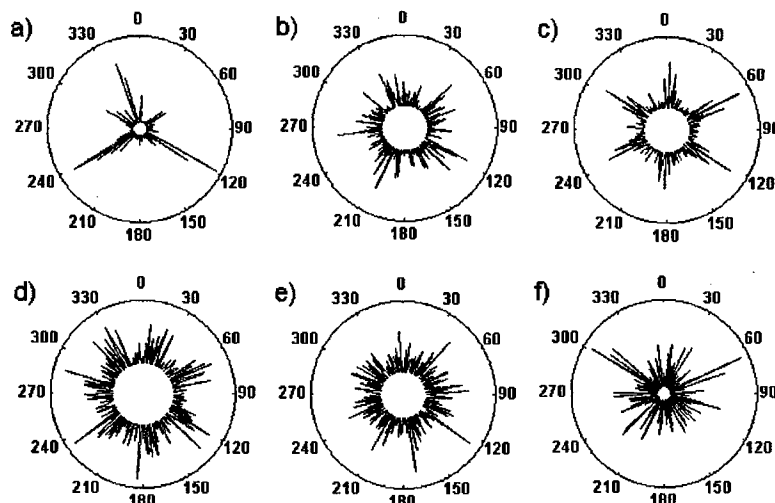


FIG. 2. Angular beams distribution for all plans of the Pareto front. The length of the radial segment in the above graphs represents the frequency of occurrence of the corresponding direction. Directions that are "preferred" during the optimization are identified with optimal directions. Presented are plans with three and five to nine beams.

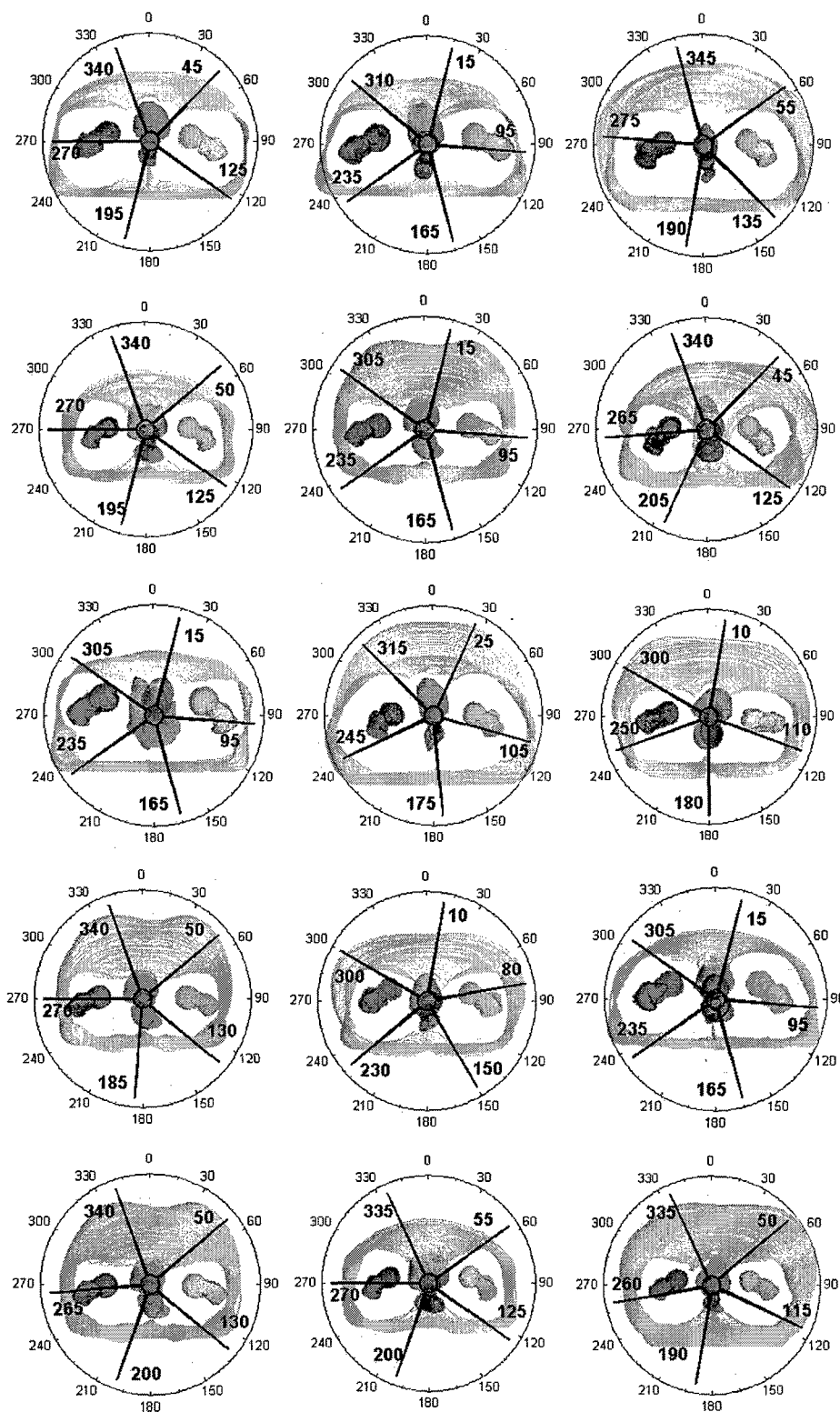


FIG. 3. The beam configuration of the most selected plan, for each individual case. Analysis of five-beam plans show that small deviations occur between plans considered optimal for each individual case. Similar results have been obtained for three and six to nine beams.

is presented in Fig. 4 (top lines). For convenience, the beam angles of the class-solution are summarized here: three beams (0° , 120° , 240°), five beams (25° , 110° , 180° , 250° , 325°), six beams (0° , 60° , 120° , 180° , 240° , 300°), seven

beams (25° , 75° , 130° , 180° , 230° , 285° , 335°), eight beams (20° , 70° , 110° , 150° , 200° , 250° , 290° , 340°), and nine beams (20° , 60° , 100° , 140° , 180° , 220° , 260° , 300° , 340°). We note that the beam angles listed here are not a "rigid"

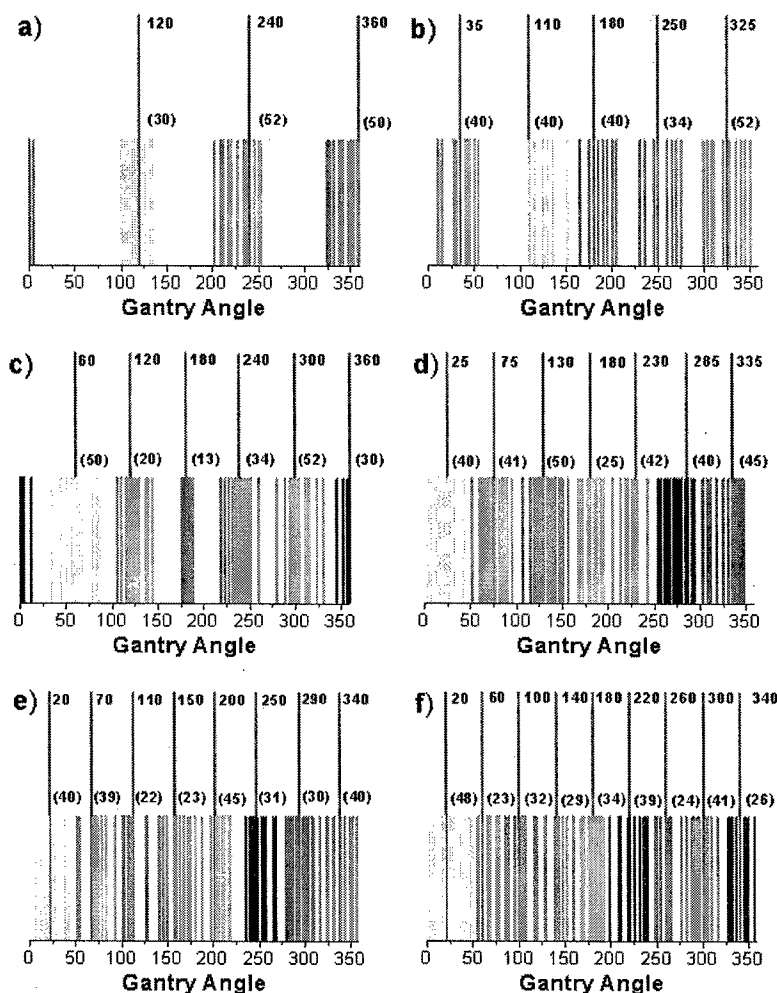


FIG. 4. Distributions of beam angles for three and five to nine beams, respectively. The shaded lines represent directions found in individual cases and the top bold lines represent the directions identified as the class-solutions. The numbers in parenthesis summarize the gantry variation ranges of the corresponding beams.

requirement and a minor adjustment of one or a few angles within their angular distributions should be acceptable in practice. The class-solution angles have been rounded to the nearest 5° angle. For three to seven beams, the locations of the maximums of the angular distributions for the fifteen patients are close to the directions of the most-selected plans. The beam setups tend to be equispaced when three to seven beams are used. We noticed that as the number of incident beams increases, the beams tend to be less confined to a few fixed directions (see Fig. 4). Instead, the optimal configuration for a case may differ from another significantly. In addition, the quality of the final plan becomes less sensitive to the beam angle variations. In other words, as the number of beams increases, the importance of optimizing beam orientation decreases. As a result, it is less necessary to derive a class-solution for IMRT prostate treatment with eight or nine beams. Practically, however, it may be desirable to have a set of fixed gantry angles to "streamline" the clinical IMRT treatment planning process. For this reason, we have

specified a set of beam configurations for these situations in Fig. 4.

C. Evaluation of class-solution approach

The validity of the class-solution approach was investigated by comparing plans obtained based on patient-specific beam orientation optimization and beam orientation class-solution. For the selected testing prostate case, two sets of plans were generated. In the first set, the optimizer uses fixed beams during the optimization process, with the directions taken from the class-solution, while in the second set, beam orientation is optimized. Both five and nine incident beams were studied.

In Fig. 5(a) we show the Pareto fronts of the two plans when five beams are used. As can be seen from the figure, the Pareto fronts obtained with the two methods are similar. For comparison, we identified two similar plans from each set and made a side-by-side comparison [Fig. 5(b)]. The two

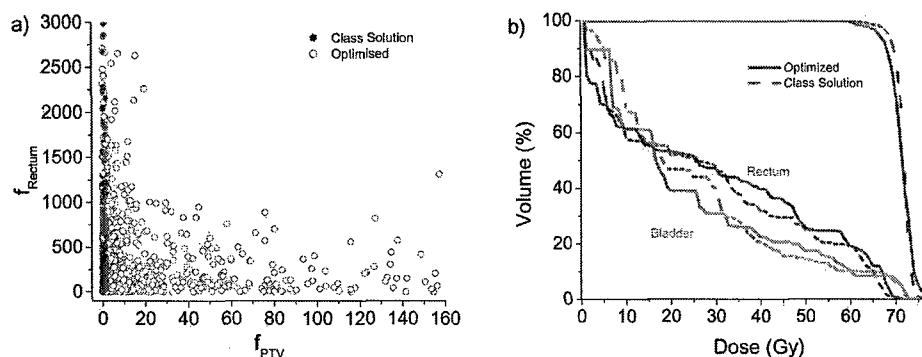


FIG. 5. Comparison of plans with optimized and class-solution plans. (a) Pareto fronts obtained by the two methods are identical for small f_{PTV} values. (b) DVHs of class-solution plan (broken line) and optimized plan show that similar plans can be obtained by the class-solution without the use of a beam optimization procedure.

plans were found to be quite similar and no clinically significant difference was found. Similar to the findings of Stein *et al.*,¹⁴ increasing the number of beams permits a better coverage of the PTV while maintaining low dose for the rectum and bladder, as reflected in the side-by-side analysis for the case of nine incident beams (Fig. 6).

We have also compared our results with that obtained using the Corvus (Nomos Corporation, Sewickly, PA), which is being used as clinical IMRT planning system at Stanford. While definitive conclusions from such a comparison is impossible because of the intrinsic differences in the two optimization systems, it seems that similar IMRT plans were obtained when the class solution angles were used. The angular solutions derived from this study were found to be consistent with our clinical observations.

IV. CONCLUSION

In general, the beam orientation class-solution is an oversimplified approach since, clinically, the anatomy, target shape and its relative location with respect to the adjacent sensitive structures, and the patient's radiation treatment history may vary from patient to patient. An individualized beam configuration is frequently, if not always, needed to

achieve the best possible treatment. In actuality, however, there exist a few disease sites (such as prostate) where the shapes, locations, and prescribed doses of various structures involved in the treatment vary slightly from case to case. In this situation, beam angle class-solution approach may be an acceptable compromise. In this work, we have studied the interpatient variation of the optimal beam configurations for prostate IMRT and derived beam orientation class-solutions for the cases of three, and five to nine incident fields based on analysis of patient-specific beam orientation optimization for fifteen prostate cases. The level of accuracy of the class-solution approach was examined by analyzing the difference between the plans obtained with class-solutions and patient specific optimizations. We found that, when the number of incident beams is less than or equal to 7, the angle variations from patient to patient are within a fairly well defined range, suggesting the possible existence of beam angle class-solution. When the number of beams is greater than seven, the class-solution becomes less meaningful but may be useful to "streamline" clinical practice. We note that in a sense any optimization algorithm is subjective due to the subjectivity in the definition of the objective function employed to model the system. It would thus be interesting to explore such model dependence and further examine the validity of

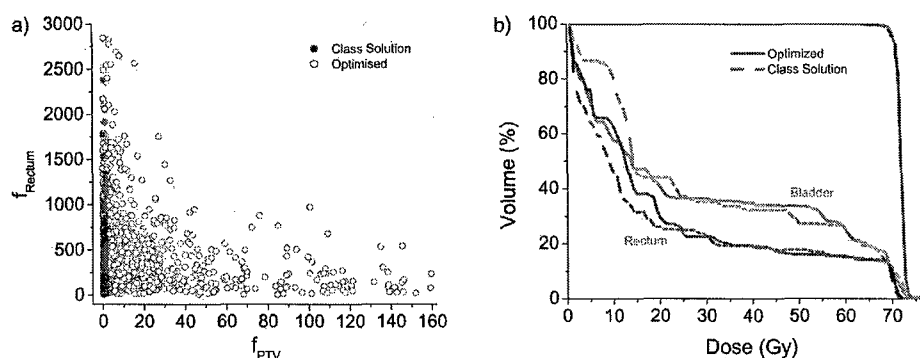


FIG. 6. Comparison of class-solution plans and optimized plans for nine beams. (a) Pareto fronts (b) DVHs. Pareto fronts obtained with class-solution directions reproduce results obtained with full optimization, concentrating solution to clinically relevant small f_{PTV} values, suggesting that a class-solution initialization for beam orientation optimization in IMRT can provide *a priori* good plans and thus reduce computation times.

the proposed class-solutions using different model systems. Given the fact that IMRT is being increasingly employed for prostate cancer irradiation, we believe that this type of study may have practical implications for clinical IMRT, especially before the users are provided with a robust beam orientation optimization tool in the clinical IMRT treatment planning systems.

ACKNOWLEDGMENTS

We would like to thank the useful discussion with Drs. S. L. Hancock, C. King, A. Boyer, J. Hai, A. Pugachev, Y. Yang, J. Lian, Z. Shou, and C. Cotrutz. This work was supported in part by a research grant from the prostate cancer research program of U.S. Department of Defense (DAMD17-03-1-0023) and the support of a Research Scholar Grant from the American Cancer Society (RSG-01-022-01-CCE).

^{a)}Author to whom correspondence should be addressed. Stanford University School of Medicine, Department of Radiation Oncology, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Electronic mail: lei@reyes.stanford.edu; Phone: (650) 498 7896; Fax: (650) 498 4015.

¹A. Pugachev and L. Xing, "Pseudo beam-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **51**(6), 1361–1370 (2001).

²L.E. Reinstein, J. Hanley, and A.G. Meek, "A feasibility study of automated inverse treatment planning for cancer of the prostate," *Phys. Med. Biol.* **41**(6), 1045–1058 (1996).

³S. Soderstrom and A. Brahme, "Which is the most suitable number of photon beam portals in coplanar radiation therapy?," *Int. J. Radiat. Oncol., Biol., Phys.* **33**(1), 151–159 (1995).

⁴P. Gokhale, E.M. Hussein, and N. Kulkarni, "Determination of beam orientation in radiotherapy planning," *Med. Phys.* **21**(3), 393–400 (1994).

⁵M.E. Hosseini-Ashrafi, H. Bagherebadian, and E. Yahaqi, "Pre-optimization of radiotherapy treatment planning: an artificial neural network classification aided technique," *Phys. Med. Biol.* **44**(6), 1513–1528 (1999).

⁶C.G. Rowbottom, M. Oldham, and S. Webb, "Constrained customization of non-coplanar beam orientations in radiotherapy of brain tumours," *Phys. Med. Biol.* **44**(2), 383–399 (1999).

⁷C.G. Rowbottom, S. Webb, and M. Oldham, "Beam-orientation customization using an artificial neural network," *Phys. Med. Biol.* **44**(9), 2251–2262 (1999).

⁸S.K. Das and L.B. Marks, "Selection of coplanar or noncoplanar beams using three-dimensional optimization based on maximum beam separation and minimized nontarget irradiation," *Int. J. Radiat. Oncol., Biol., Phys.* **38**(3), 643–655 (1997).

⁹M. Braunstein and R.Y. Levine, "Optimum beam configurations in tomographic intensity modulated radiation therapy," *Phys. Med. Biol.* **45**(2), 305–328 (2000).

¹⁰A. Pugachev, A. Boyer, and L. Xing, "Beam orientation optimization in intensity-modulated radiation treatment planning," *Med. Phys.* **27**(6), 1238–1245 (2000).

¹¹G.A. Ezzell, "Genetic and geometric optimization of three-dimensional radiation therapy treatment planning," *Med. Phys.* **23**(3), 293–305 (1996).

¹²S.L. Sailer, J.G. Rosenman, J.R. Symon, T.J. Cullip, and E.L. Chaney, "The tetrad and hexad: maximum beam separation as a starting point for noncoplanar 3D treatment planning: prostate cancer as a test case," *Int. J. Radiat. Oncol., Biol., Phys.* **30**(2), 439–446 (1994).

¹³J.L. Bedford and S. Webb, "Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy," *Med. Phys.* **30**(7), 1788–1804 (2003).

¹⁴J. Stein, R. Mohan, X.H. Wang, T. Bortfeld, Q. Wu, K. Preiser, C.C. Ling, and W. Schlegel, "Number and orientations of beams in intensity-modulated radiation treatments," *Med. Phys.* **24**(2), 149–160 (1997).

¹⁵M. Asell, S. Hyodynmaa, S. Soderstrom, and A. Brahme, "Optimal electron and combined electron and photon therapy in the phase space of complication-free cure," *Phys. Med. Biol.* **44**(1), 235–252 (1999).

¹⁶O.C. Haas, K.J. Burnham, and J.A. Mills, "Optimization of beam orientation in radiotherapy using planar geometry," *Phys. Med. Biol.* **43**(8), 2179–2193 (1998).

¹⁷T. Bortfeld and W. Schlegel, "Optimization of beam orientations in radiation therapy: some theoretical considerations," *Phys. Med. Biol.* **38**(2), 291–304 (1993).

¹⁸A. Pugachev and L. Xing, "Incorporating prior knowledge into beam orientation optimization," *Int. J. Radiat. Oncol., Biol., Phys.* **54**, 1565–1574 (2002).

¹⁹A.B. Pugachev, A.L. Boyer, and L. Xing, "Beam orientation optimization in intensity-modulated radiation treatment planning," *Med. Phys.* **27**(6), 1238–1245 (2000).

²⁰S. Das, T. Cullip, G. Tracton, S. Chang, L. Marks, M. Anscher, and J. Rosenman, "Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization," *Int. J. Radiat. Oncol., Biol., Phys.* **55**(1), 215–224 (2003).

²¹D. Djajaputra, Q. Wu, Y. Wu, and R. Mohan, "Algorithm and performance of a clinical IMRT beam-angle optimization system," *Phys. Med. Biol.* **48**, 3191–3212 (2003).

²²G. Meedt, M. Alber, and F. Nüsslin, "Non-coplanar beam direction optimization for intensity-modulated radiotherapy," *Phys. Med. Biol.* **48**, 2999–3019 (2003).

²³C.M. Nutting, C.G. Rowbottom, V.P. Cosgrove, J.M. Henk, D.P. Dearnaley, M.H. Robinson, J. Conway, and S. Webb, "Optimisation of radiotherapy for carcinoma of the parotid gland: a comparison of conventional, three-dimensional conformal, and intensity-modulated techniques," *Radiother. Oncol.* **60**(2), 163–172 (2001).

²⁴G.A. Ezzell, S.E. Schild, and W.W. Wong, "Development of a treatment planning protocol for prostate treatments using intensity modulated radiotherapy," *J. Appl. Clin. Med. Phys.* **2**(2), 59–68 (2001).

²⁵J.H. Mott, J.E. Livsey, and J.P. Logue, "Development of a simultaneous boost IMRT class solution for a hypofractionated prostate cancer protocol," *Br. J. Radiol.* **91**(77), 377–386 (2004).

²⁶L. Xing, J.G. Li, S. Donaldson, Q.T. Le, and A.L. Boyer, "Optimization of importance factors in inverse planning," *Phys. Med. Biol.* **44**(10), 2525–2536 (1999).

²⁷Y. Yu, "Multiobjective decision theory for computational optimization in radiation therapy," *Med. Phys.* **24**(4), 1445–1454 (1997).

²⁸E. Schreibmann, M. Lahanas, L. Xing, and D. Baltas, "Multiobjective evolutionary optimization of number of beams, their orientation and weights for IMRT," *Phys. Med. Biol.* **49**(5), 747–770 (2004).

²⁹Q. Hou, J. Wang, Y. Chen, and J.M. Galvin, "Beam orientation optimization for IMRT by a hybrid method of the genetic algorithm and the simulated dynamics," *Med. Phys.* **30**(30), 2271–2567 (2003).

³⁰K. Deb and M. Goyal, "A combined genetic adaptive search (GeneAS) for engineering design," *Comp. Sci. Inform.* **26**, 30–45 (1996).

³¹D.C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.* **45**, 503–528 (1989).

³²M. Lahanas, E. Schreibmann, N. Milickovic, and D. Baltas, "Intensity modulated beam radiation therapy dose optimization with multiobjective evolutionary algorithms," *Lect. Notes Comput. Sci.* **2632**, 648–661 (2003).

³³M. Lahanas, D. Baltas, S. Giannouli, N. Milickovic, and N. Zamboglou, "Generation of uniformly distributed dose points for anatomy-based three-dimensional dose optimization methods in brachytherapy," *Med. Phys.* **27**, 1034–1046 (2000).

³⁴L.J. Boersma, M. van den Brink, A.M. Bruce, T. Shouman, L. Gras, A. te Velde, and J.V. Lebesque, "Estimation of the incidence of late bladder and rectum complications after high-dose (70–78 Gy) conformal radiotherapy for prostate cancer, using dose-volume histograms," *Int. J. Radiat. Oncol., Biol., Phys.* **41**(1), 83–92 (1998).

³⁵B. Emami, J. Lyman, A. Brown, L. Coia, M. Goitein, J.E. Munzenrider, B. Shank, L.J. Solin, and M. Wesson, "Tolerance of normal tissue to therapeutic irradiation," *Int. J. Radiat. Oncol., Biol., Phys.* **21**(1), 109–122 (1991).

³⁶C. Cotrutz, M. Lahanas, C. Kappas, and D. Baltas, "A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy," *Phys. Med. Biol.* **46**(8), 2161–2175 (2001).

Clinical knowledge-based inverse treatment planning

Yong Yang and Lei Xing

Department of Radiation Oncology, Stanford University School of Medicine, Stanford,
CA 94305-5847, USA

E-mail: lei@reyes.stanford.edu

Received 1 June 2004, in final form 20 September 2004

Published 25 October 2004

Online at stacks.iop.org/PMB/49/5101

doi:10.1088/0031-9155/49/22/006

Abstract

Clinical IMRT treatment plans are currently made using dose-based optimization algorithms, which do not consider the nonlinear dose–volume effects for tumours and normal structures. The choice of structure specific importance factors represents an additional degree of freedom of the system and makes rigorous optimization intractable. The purpose of this work is to circumvent the two problems by developing a biologically more sensible yet clinically practical inverse planning framework. To implement this, the dose–volume status of a structure was characterized by using the effective volume in the voxel domain. A new objective function was constructed with the incorporation of the volumetric information of the system so that the figure of merit of a given IMRT plan depends not only on the dose deviation from the desired distribution but also the dose–volume status of the involved organs. The conventional importance factor of an organ was written into a product of two components: (i) a generic importance that parametrizes the relative importance of the organs in the ideal situation when the goals for all the organs are met; (ii) a dose-dependent factor that quantifies our level of clinical/dosimetric satisfaction for a given plan. The generic importance can be determined *a priori*, and in most circumstances, does not need adjustment, whereas the second one, which is responsible for the intractable behaviour of the trade-off seen in conventional inverse planning, was determined automatically. An inverse planning module based on the proposed formalism was implemented and applied to a prostate case and a head–neck case. A comparison with the conventional inverse planning technique indicated that, for the same target dose coverage, the critical structure sparing was substantially improved for both cases. The incorporation of clinical knowledge allows us to obtain better IMRT plans and makes it possible to auto-select the importance factors, greatly facilitating the inverse planning process. The new formalism proposed also reveals the relationship between different inverse planning schemes and gives important insight into the problem of therapeutic plan optimization.

In particular, we show that the EUD-based optimization is a special case of the general inverse planning formalism described in this paper.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

An important issue in inverse planning is how to formalize the clinical goals to objectively evaluate the figures of merit of different IMRT plans. Despite intense research effort in modelling the clinical decision-making strategies (Amols and Ling 2002, Deasy *et al* 2002, Earl *et al* 2003, Hou *et al* 2003, Lahanas *et al* 2003, Langer *et al* 1993, 1998, Lee *et al* 2000, Llacer *et al* 2001, Mohan *et al* 1994, Webb 2004, Xing *et al* 1999, Yan *et al* 2003), the appropriate form of the objective function remains illusive. Presently, two types of objective functions are widely used: dose or dose-volume histogram (DVH)-based (physical objective functions) (Chen *et al* 2002, Cho *et al* 1998, Holmes *et al* 1995, Hristov *et al* 2002, Michalski *et al* 2004, Starkschall *et al* 2001, Shepard *et al* 2002, Xing *et al* 1998) and dose-response-based objective functions (biological objective functions) (Brahme 2001, Kallman *et al* 1992, Miften *et al* 2004, Mohan *et al* 1992, Wang *et al* 1995, Webb and Nahum 1993). The underlying difference between these models lies in what endpoints are used to evaluate the treatment or which fundamental quantities are used to define the optimality. The physical approach emphasizes the difference between the calculated and prescribed doses and does not consider the nonlinear effects for tumours and normal structures. Dose-volume constraints are often introduced to select a solution with certain shapes of the DVHs for the target and sensitive structures. However, it is important to note that the construction of the DVH constraints is *a priori* in nature. The use of constraints can only passively restrict the accessible DVHs by narrowing the solution space, and the figures of merits of the physically realizable plans are not changed as long as they satisfy the constraints. To reflect our preference over certain DVHs for a structure, it is necessary to express the objective of the structure as a function of the volumetric status, which has not been achieved up to this point. On the other hand, biological model-based optimization proponents argue that plan optimization should be guided by estimates of biological effects, which depend on the radiation dose through the dose-response function. The treatment objective in biological model-based inverse planning is usually stated as the maximization of the tumour control probability (TCP) while maintaining the normal tissue complication probability (NTCP) to within acceptable levels (Brahme 2001, Kallman *et al* 1992, Langer *et al* 1998, Mohan *et al* 1992, Wang *et al* 1995). In principle, the biologically based models are most relevant for radiotherapy plan ranking. However, the dose-response function of various structures is not sufficiently understood and, at this point, there is considerable controversy about the models for computing dose-response indices and their use in optimization. In reality, the use of dose-response indices for optimization may lead to very inhomogeneous target dose distributions. Furthermore, it is difficult for clinicians to specify the optimization criteria in terms of dose-response indices. This becomes compounded when two or more independently optimized plans are to be combined. Because of the problems, the use of biological model-based optimization has mainly been limited in the research community and little effort has been made to implement the model in commercial planning systems. Given the fact that biological outcome is the ultimate endpoint of radiation therapy, the importance of the biological modelling and the biological model-based inverse planning can never be overstated.

To pin down the underlying problem of the current inverse planning formalism and illustrate the need for a clinically more relevant approach, let us take parotid glands as an example. It is well known that the clinical endpoint is the same if the glands are irradiated with 15 Gy to 67%, or 30 Gy to 45%, or 45 Gy to 24% of the total volume (Eisbruch *et al* 1999). If a dose-based metric, such as the commonly used quadratic objective function, is used, the rankings for the three different scenarios would be completely different. Even with the use of dose-volume constraints, it is difficult, if not impossible, to incorporate this type of knowledge to correctly model the behaviour of the organ in response to radiation. Indeed, a constraint in optimization acts as a 'boundary condition' during the optimization and does not change the rankings of dosimetrically different plans. This example clearly reveals the inadequacy of the conventional dose-based objective function and suggests the urgent need for a clinically more sensible model. Obviously, a minimum requirement for the model is that it should be consistent with the existing clinical outcome data. For parotid glands, for instance, the three different DVHs mentioned above should be scored equally. This type of 'degeneracy' can be achieved by effectively integrating clinical endpoint data into the inverse planning formulation. For a given patient, the specific DVH selection will be determined by the optimization algorithm with the consideration of the dosimetric/clinical requirements of other structures and the results will, of course, depend on the geometric and dosimetric details of the given patient. The example, however, underscores the important role of the existing clinical data in inverse planning and emphasizes the essential ingredients for a clinically realistic objective function. Towards developing a biologically more sensible yet clinically practical inverse planning formalism, in the following we propose a method to incorporate clinical endpoint data into the construction of the objective function and attempt to bridge the gap between the clinical decision-making process and the computational modelling. Our study indicates that the clinical knowledge-based modelling allows us to objectively rank IMRT plans according to their clinical merits and makes it possible to obtain truly optimal IMRT plans with much reduced efforts.

2. Methods and materials

2.1. Dose-volume effect

Generally, the dose response of a structure with respect to the irradiated dose and volume is complicated. The fact that the dose distribution in tumour or a sensitive structure is generally inhomogeneous makes the establishment of such a relationship even more intractable. Over the last two decades, attempts have been made by many researchers to capture the main feature(s) of the dose-volume effects. A power-law model represents one of the successful techniques in dealing with the dose-volume effects of sensitive structures (Lyman and Wolbarst 1987). In this model an equivalent dose uniformly irradiating the whole organ, D_{eq} , can be used to represent the situation in which a fractional partial volume, v , is irradiated to a dose, D , by a simple power-law model: $D_{eq} = v^{1/n} D$. A remarkable characteristic of this model is that, although only a single organ-specific parameter, n , is used, clinical and biological data have shown that this power law holds well at low complication levels (Lyman and Wolbarst 1987, Schultheiss *et al* 1983). On the basis of this relation, Mohan *et al* (1992) introduced the concept of effective dose to represent a non-uniform dose distribution in a sensitive structure. Kutcher and Burman (1989) applied the same power model independently to each volume element of the histogram and introduced the concept of effective volume to reduce the DVH of an inhomogeneous dose distribution in a sensitive structure to a uniform dose distribution.

Following their study, in this work we define the effective volume $(\Delta V_{\text{eff}})_i$ for a voxel i with volume ΔV and dose $D_c(i)$ as follows:

$$(\Delta V_{\text{eff}})_i = \Delta V (D_c(i)/D_{\text{ref}})^{1/n} \quad (1)$$

and extend this concept to handle the voxels in the tumour target, where n is an organ-dependent parameter, and D_{ref} is the reference dose. For a sensitive structure, n is a small positive number ($0 < n < 1$) and the value of parameter n reflects the architecture (serial or parallel) of the sensitive structure. For a target, n should be assigned with a small negative value ($-1 < n < 0$). The biological meaning of equation (1) is that for a sensitive structure a small volume receiving a higher dose than a reference dose would be equivalent to a larger volume receiving the reference dose; for a target, a volume with a lower dose would have a larger effective volume. The effective volumes of all voxels reflect the DVH status of the given organ, and for inverse planning, this permits us to deal with the complicated dose-volume effect in the voxel domain.

2.2. Dose-volume-based objective function

The objective function, f , expressed as a function of the effective volume in the voxel domain for an organ should take the form of

$$f = f(\{(\Delta V_{\text{eff}})_i\}). \quad (2)$$

Generally, the dose-volume effect suggests that the voxels receiving different doses are inequivalent: the one with a larger effective volume (higher dose for a critical organ) should be penalized more when compared to a voxel with a smaller effective volume (lower dose). Thus we heuristically write the f in the following form:

$$f = 1 + \eta_1 \sum_i r_i [D_c(i)/D_{\text{ref}}]^{1/n} + \eta_2 \left\{ \sum_i r_i [D_c(i)/D_{\text{ref}}]^{1/n} \right\}^2 + \dots, \quad (3)$$

where equation (1) has been incorporated, r_i is the importance factor of the i th voxel, representing the intrastructural trade-off due to physical/clinical requirements other than the dose-volume-based penalty, η_1 and η_2 are phenomenological parameters of the model. In equation (3), the third (and higher order) term emphasizes more the voxels with high effective volumes, whereas the first and second terms ensure that the voxels with low effective volumes receive an adequate penalty. We typically set $r_i \equiv 1$, unless there are other physical/clinical considerations (e.g., when the density of clonogens varies spatially (Xing *et al* 2002)). In this work, we set $\eta_1 = 1$ and $\eta_2 = \eta_3 = \dots = 0$.

2.3. Hybrid of dose-based and dose-volume-based objective functions

Equation (3) provides a good description of the dose-volume effect. With proper choice of the parameter, n , the clinically observed dose-volume effect can be reproduced by the objective function. In reality, other requirements, such as the target dose uniformity, should also be considered. A more general form of inverse planning objective function can be written as a hybrid of the dose-volume-based and the dose-based functions. In this situation, the overall objective function of the system takes the form of

$$F = \sum_{\tau=1}^{t_{\tau}} r_{\tau} \frac{1}{N_{\tau}} \sum_{i=1}^{N_{\tau}} \left\{ 1 + \eta'_{\tau} [D_c(i)/D_{\tau,\text{ref}}]^{1/n_{\tau}} \right\} |D_c(i) - D_0^T(i)|^{k_{\tau}} \\ + \sum_{\sigma=1}^{s_{\sigma}} r_{\sigma} \frac{1}{N_{\sigma}} \sum_{i=1}^{N_{\sigma}} \left\{ 1 + \eta^s_{\sigma} [D_c(i)/D_{\sigma,\text{ref}}]^{1/n_{\sigma}} \right\} D_c(i)^{k_{\sigma}}, \quad (4)$$

where t_τ and s_σ are the numbers of targets and sensitive structures, $D_0^T(i)$ is the prescription dose in target voxel i , subscripts τ and σ represent target τ and sensitive structure σ , N_τ , N_σ , r_τ , r_σ , n_τ , n_σ , $D_{\tau,\text{ref}}$, $D_{\sigma,\text{ref}}$, k_τ and k_σ represent the total numbers of voxels, structure specific importance factors, n parameters, reference doses, power of dosimetric deviation from the specified criteria for target τ and sensitive structure σ , respectively. The factor $|D_c(i) - D_0^T(i)|^{k_\tau}$ for target or $D_c(i)^{k_\sigma}$ for a sensitive structure represents the contribution from dosimetric deviation from the ideal situation. If the k_τ and k_σ are set to zero, the objective function becomes purely dose-volume driven. In particular, if we set k_σ to zero and k_τ to a nonzero value, the objective function for a target becomes a hybrid of dose-volume and dose, whereas the objective functions for critical structures remain purely dose-volume based. On the other hand, when all the n parameters in equation (4) are set to be $+\infty$, no dose-volume effects are considered and equation (4) is reduced to the conventional dose-based objective function.

2.4. Automatic determination of structure specific importance factors

The selection of structure specific importance factors, r_τ or r_σ in equation (4), is generally done empirically by trial and error. Here we describe an automated approach for solving the problem. The key to success is to establish an effective method to express the structural importance factor in terms of physically or clinically more meaningful quantities. For this purpose, we write the importance of a structure into a product of two components: (i) a generic factor parametrizing the relative importance of the organs in an ideal situation when the goals for the organs are met; and (ii) a dose-dependent factor quantifying our level of clinical/dosimetric satisfaction for a given plan. The first factor can be determined *a priori*, and in most circumstances, does not need adjustment (generally speaking, the value of r_σ^g is determined on the basis of the treatment modality and the patient's overall condition), whereas the second one is responsible for the intractable behaviour of the trade-off in conventional planning and can be automatically determined. This decomposition is essentially to normalize the conventional importance factor in terms of our clinical goals for the structures under discussion. Because of this decomposition, the meaning of the importance factor becomes more transparent and the determination of the factors becomes straightforward. Mathematically, we write $r_\sigma = r_\sigma^g r_\sigma^d$, where r_σ^g represents the first contribution described above (the desired weighting among different structures in an ideal situation), and r_σ^d is the second component and is defined as a function of NTCP for a sensitive structure. In this study r_σ^g was set empirically (see tables 2 and 3 for examples). r_σ^d is updated according to the DVH or the dose distribution during the optimization process and reflects the most current status of trade-off in the system. Generally, the importance of a sensitive structure should be increased in the next iteration if NTCP is high, and *vice versa*. We found that a simple linear relation between r_σ^d and NTCP,

$$r_\sigma^d = \text{NTCP}_\sigma + \delta, \quad (5)$$

describes the trade-off behaviour of the system well, where the value of NTCP_σ depends on the dose distribution at the current iteration for structure σ , and δ is a cut-off factor for NTCP, which is introduced to ensure the sensitive structure receives a minimum penalty even if its NTCP is close to zero. We set δ as an organ-independent constant of 0.01%.

The NTCP was assessed using Lyman's model in this study. For non-uniform irradiation, the Kutcher-Burman (1989) effective-volume DVH reduction method is used to transform a DVH into a uniform irradiation on an effective partial volume. Model parameters (n , m ,

TD_{50/5}) used in this study were those fitted to the model by Burman *et al* (1991) for the normal tissue tolerance data compiled by Emami *et al* (1991).

2.5. Computational algorithm

After considering the automatic trade-off strategy, the generalized objective function takes the following form:

$$F = \sum_{\tau=1}^{t_{\tau}} r_{\tau} \frac{1}{N_{\tau}} \sum_{i=1}^{N_{\tau}} \left\{ 1 + \eta_1^t [D_c(i)/D_{\tau,ref}]^{1/n_{\tau}} \right\} |D_c(i) - D_0^T(i)|^{k_{\tau}} \\ + \sum_{\sigma=1}^{s_{\sigma}} r_{\sigma}^g r_{\sigma}^d \frac{1}{N_{\sigma}} \sum_{i=1}^{N_{\sigma}} \left\{ 1 + \eta_1^s [D_c(i)/D_{\sigma,ref}]^{1/n_{\sigma}} \right\} D_c(i)^{k_{\sigma}}. \quad (6)$$

We implemented a software module to optimize the objective function (6) in the PLUNC treatment planning system (University of North Carolina, Chapel Hill, NC). The ray-by-ray iterative algorithm (SIITP) reported earlier (Xing *et al* 1998) was employed to obtain the optimal beam intensity profiles. The values of k_{τ} and k_{σ} in equation (6) were set to be 2, but the behaviour of the system for a few other combinations of k_{τ} and k_{σ} were also checked for the prostate case. The reference dose, $D_{\sigma,ref}$, was chosen to be TD_{5/5} of the corresponding critical organ. For the target, $D_{\tau,ref}$ was set as the prescription dose. Figure 1 shows the flow chart of the calculation process. In the current study we specify a maximum number of iterations as the termination condition of the optimization process. The DVHs can be inspected in each iterative step to visually monitor the optimization process.

2.6. Case studies

Two cases, a prostate case and a head-neck case, were used to evaluate the proposed inverse planning formalism. The optimization results were compared with those obtained using the conventional dose-based optimization method, which was described in detail by Xing *et al* (1998). The optimization parameters in the dose-based method were adjusted by trial-and-error to obtain an 'optimal' plan.

In the prostate case, the target volume included the prostate and seminal vesicles. The sensitive structures included rectum, bladder and femoral heads. All the IMRT plans used identical configuration of five equally spaced 15 MV photon beams with gantry angles of 0°, 72°, 144°, 216° and 288° (in IEC convention). The plans were normalized to deliver the prescription dose of 70 Gy to 99% of the target volume. The parameter n_{τ} was chosen to be -0.2 for the target. The parameters used in the NTCP calculations of the rectum, bladder and femoral heads are listed in table 1. Table 2 summarizes the optimization parameters for both the newly proposed and dose-based approaches.

For the prostate case, we also studied the influence of two more combinations of k_{σ} and k_{τ} . These included ($k_{\sigma} = 2, k_{\tau} = 4$) and ($k_{\sigma} = 0, k_{\tau} = 2$). In the latter case, we have included a higher order term of the dose-volume effect (the third term in equation (3) with $\eta_2 = 1$) to ensure that the high effective volume voxels are penalized enough in the absence of the dose-based factor.

In the head-and-neck case, the organs at risk included the eyes, optic nerves, optic chiasm, brainstem, spinal cord and parotids. Two targets were the gross target volume (GTV) and the clinical target volume (CTV), which includes the microscopic disease region surrounding the GTV. The plan was normalized to deliver a prescription dose of 70 Gy to at least 99% of the GTV and 62 Gy to at least 95% of the CTV. Nine equally spaced 6 MV coplanar beams

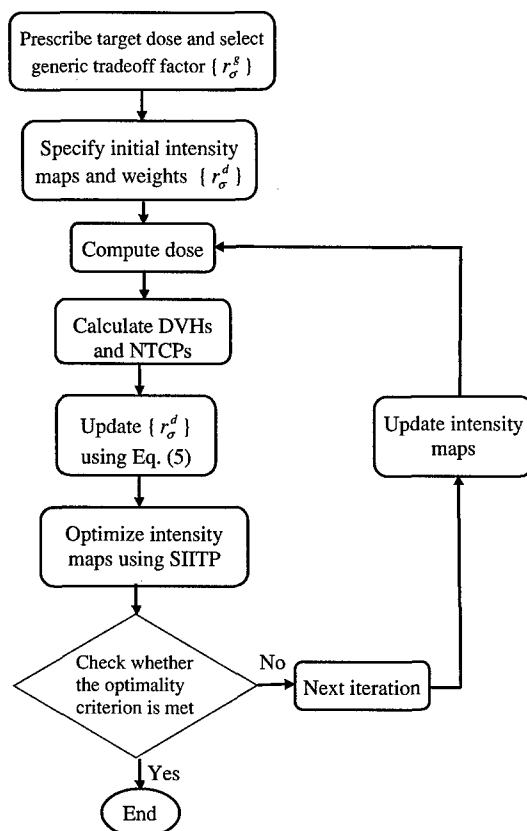


Figure 1. A flow chart of the proposed optimization process.

Table 1. The radiological parameters for various sensitive structures used in this study.

Sensitive structures	n	m	$D_{50/5}$ (Gy)	$D_{5/5}$ (Gy)
Bladder	0.50	0.11	80	65
Rectum	0.12	0.15	80	60
Femoral head	0.25	0.12	65	52
Eye lens	0.3	0.27	18	10
Optic nerve	0.25	0.14	65	50
Optic chiasm	0.25	0.14	65	50
Spinal cord	0.05	0.175	66.5	47
Brainstem	0.16	0.14	65	50
Parotid	0.70	0.18	46	32

(0°, 40°, 80°, 120°, 160°, 200°, 240°, 280° and 320°) were used for this case. The parameter n_{τ} was -0.5 for both GTV and CTV. The parameters used for the computation of the NTCPs of the sensitive structures are also obtained from the same source stated earlier and are listed in table 1. The optimization parameters for both the two techniques are summarized in table 3.

Table 2. Summary of the optimization parameters used in the dose-based and proposed approaches for the prostate case.

Organs	The dose-based approach		The proposed approach Generic importance factors (r_g^g)
	Relative importance factors	Target prescription and OAR tolerance doses (Gy)	
Target	5.0	78	5
Bladder	1.2	48	2
Rectum	1.8	43	2
Femoral head (R)	1.0	32	1
Femoral head (L)	1.0	32	1
Normal tissue	0.5	65	0.3

Table 3. Summary of the optimization parameters used in the dose-based and proposed approaches for the head-and-neck case.

Organs	The dose-based approach		The proposed approach Generic importance factors (r_g^g)
	Relative importance factors	Target prescription and OAR tolerance doses (Gy)	
GTV	3.0	70	4.0
CTV	4.0	62	6.0
Spinal cord	2.0	30	3.0
Brainstem	1.5	30	2.0
Left optic nerve	1.0	25	1.0
Right optic nerve	1.0	25	1.0
Left eye	2.0	6	3.0
Right eye	2.0	6	3.0
Left parotid	1.2	25	1.0
Right parotid	1.2	25	1.0
Optic chiasm	1.0	25	1.0
Normal tissue	0.5	40	0.5

3. Results

3.1. Prostate IMRT plans

Figures 2 and 3 summarize the results of the two IMRT plans obtained using the newly proposed and conventional techniques. Figure 2 compares the isodose distributions in two transverse slices and a sagittal slice for the two plans. The DVHs of the target and sensitive structures are plotted in figure 3, in which the solid and dashed lines represent the DVHs obtained using the new and conventional approaches, respectively. The calculated NTCPs of rectum, bladder and femoral heads for both IMRT plans are listed in table 4. According to the table, it is seen that the NTCPs of the sensitive structures are improved significantly. For the rectum, for example, the NTCP is reduced from 0.45% to 0.03%. Our results also indicate that the main compromise in a prostate IMRT treatment seems to be between the tumour coverage and the rectum complication because the NTCP of the rectum is much higher than that of other sensitive structures.

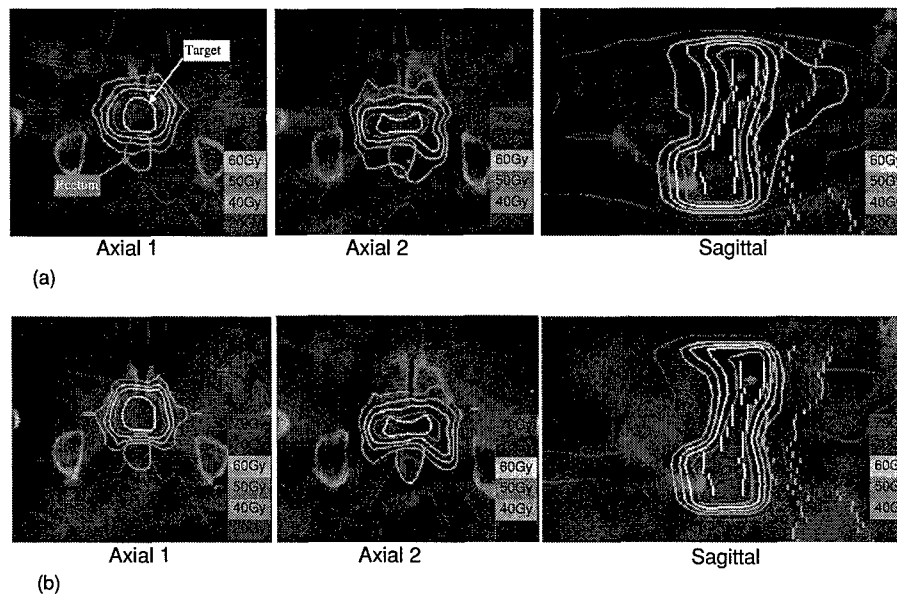


Figure 2. Comparison of the isodose distributions of the two prostate IMRT plans: (a) the conventional dose-based approach; (b) the newly proposed approach.

Table 4. Comparison of the NTCP for the two IMRT plans for the prostate case.

NTCP (%)	The dose-based IMRT plan	The proposed IMRT plan
Bladder	0.017	0.000 30
Rectum	0.45	0.029
Femoral head (R)	0.000 076	0.000 0038
Femoral head (L)	0.000 032	0.000 015

The above results demonstrate that, for comparable target coverage, the new inverse planning technique greatly improves the critical structure sparing, especially the rectum sparing. By comparing the isodose distributions of the two plans (figure 2), it is seen that the dose gradient at the interface between the target and the rectum is much steeper for the IMRT plan obtained with the new formalism. Furthermore, it is intriguing that the non-sensitive structure normal tissue also receives fewer doses in comparison with that of the dose-based optimization. Our results suggest that the improvement in the critical structure sparing is achieved not at the cost of higher target dose inhomogeneity, which is commonly seen in IMRT plan optimization.

The resultant DVHs when $k_\sigma = 2$ and k_τ was increased from 2 to 4 in equation (6) are plotted in figure 3 as the dotted curves. While the target dose uniformity is improved when the k_τ increases, the doses to the rectum and bladder are worsened. The results make intuitive sense as when the k_τ increases, more penalty is applied towards dosimetric deviation from the prescription. The DVHs when the objective function for the target is a hybrid of dose-volume and dose-based functions ($k_\tau = 2$) and that for the sensitive structures are purely dose-volume based ($k_\sigma = 0$) are shown in figure 3 as dash-dotted curves. In this case, a high order term

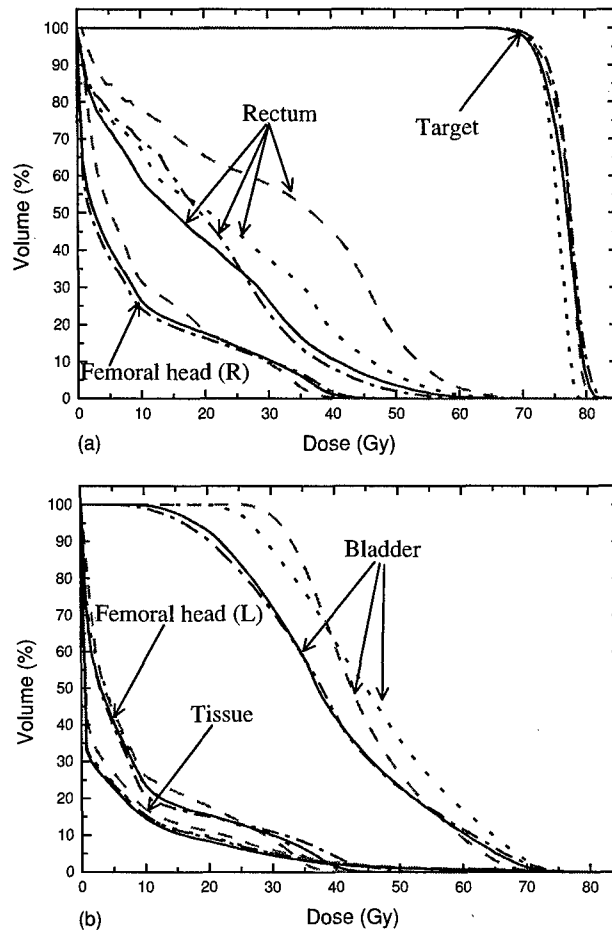


Figure 3. Comparison of DVHs of the prostate IMRT plans obtained using the proposed approach (solid curves) and the conventional dose-based approach (dashed curves). The dotted curves represent the results obtained with $k_t = 4$ and $k_\sigma = 2$ in equation (6). The dash-dotted curves are the DVHs with $k_t = 2$ and $k_\sigma = 0$ (a higher order term, the third term in equation (3), was included during the optimization).

of the dose-volume effect in equation (3) was added to ensure that the high effective volume voxels are penalized enough in the absence of the dose-based factor. Interestingly, as can be seen from figure 3, the results so obtained were very similar to that obtained with the hybrid objective function.

3.2. Head-and-neck IMRT plans

Figure 4 shows the isodose distributions obtained using the two different planning techniques in three transverse slices, one sagittal slice and one coronal slice for the two plans. Figure 5 compares the DVHs of the targets and sensitive structures, in which the solid and dashed lines represent the DVHs obtained using the newly proposed and conventional approaches,

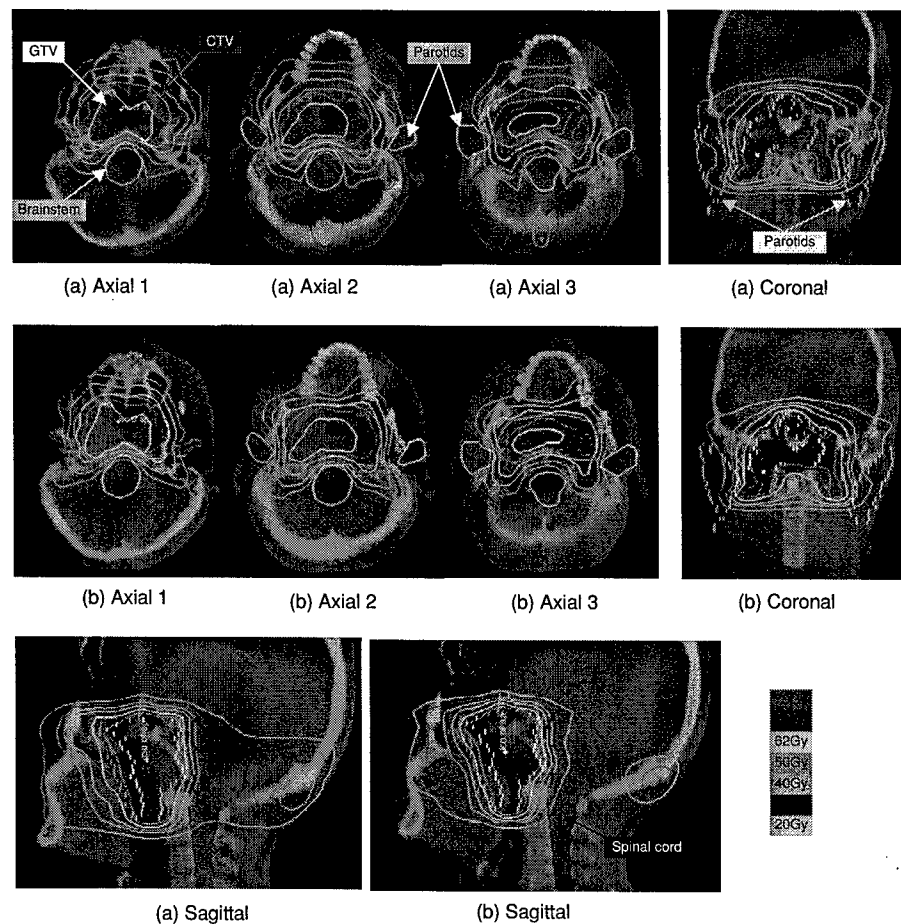


Figure 4. Comparison of the isodose distributions of the two head-and-neck IMRT plans: (a) the conventional dose-based approach; (b) the newly proposed approach.

Table 5. Comparison of the NTCP for the two IMRT plans for the head-and-neck case.

NTCP (%)	The dose-based IMRT plan	The proposed IMRT plan
Spinal cord	0.043	0.0025
Brainstem	0.012	0.0040
Left Eye	0.27	0.18
Right Eye	0.24	0.12
Left parotid	0.21	0.056
Right parotid	0.22	0.064
Optic chiasm	0.00024	0.000064
Left optic nerve	0.000064	0.0000075
Right optic nerve	0.000043	0.000025

respectively. The calculated NTCPs of eyes, optical nerves, optical chiasm, brainstem, spinal cord and parotids for both plans are shown in table 5. As seen from the isodose distributions

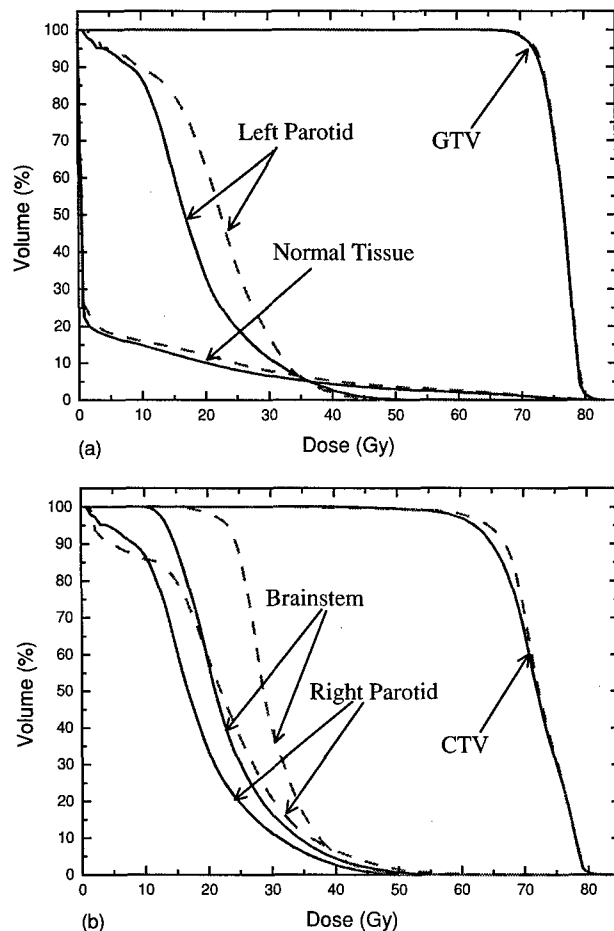


Figure 5. Comparison of the DVHs of the two head-and-neck IMRT plans obtained using our newly proposed approach (solid curves) and the conventional dose-based approach (dashed curves).

(figure 4) and DVHs (figure 5), with comparable GTV and CTV dose coverage and dose homogeneity, the doses to the sensitive structures are dramatically reduced. The dose reduction is particularly pronounced in the spinal cord, brainstem, parotids and eyes. For the left and right parotids, for example, the fractional volume receiving a dose above 25 Gy is reduced from 35% to 20% and 15%, respectively. Consistent with the enhanced dosimetric conformality and similar to the prostate case, much steeper dose gradient occurs near the boundary of the target volume. The dose to the non-sensitive structure normal tissue is also lower in comparison with the conventional IMRT plan. While the NTCPs of the optical nerves and optical chiasm are small and it is difficult to draw a conclusion, from table 5, it is quite clear that the NTCPs of the eyes, parotids, spinal cord and brainstem are improved significantly. We emphasize once again that the significant improvement in sensitive structure sparing is achieved without deteriorating the dose coverage of the GTV and CTV.

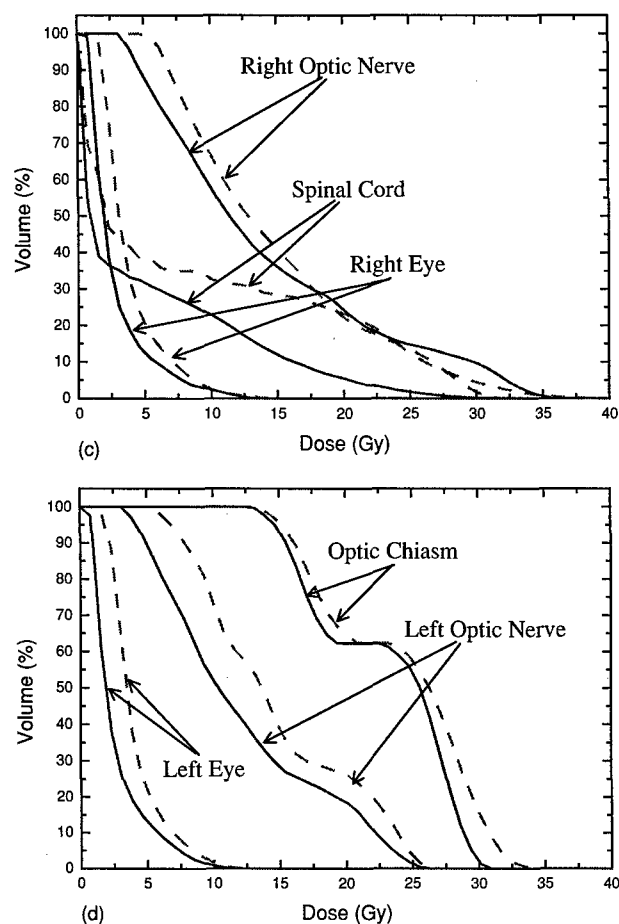


Figure 5. (Continued.)

4. Discussion

The currently available dose-based objective functions do not truly reflect the nonlinear relationship between the dose and the response of tumours and tissues, and it is highly desirable to incorporate clinical outcome data in the formulation of inverse planning to guide the plan optimization process. While the dose dependence of a clinical endpoint may be degenerate in the sense that it may be caused by a variety of dose distributions or DVHs, there exists no mechanism in conventional inverse planning to model the phenomenon. The irradiation of parotid glands mentioned in the introduction represents an example of this. On the other hand, the conventional objective function may impose some unrealistic degeneracy that is inconsistent with clinical experience. For example, assume that in a treatment plan the prescription dose to the target is 70 Gy and that the tumour is divided into two parts with the same volume, one receiving a dose 60 Gy and another 80 Gy. The penalty values of the two scenarios would be the same according to the conventional quadratic function. Obviously, the

two different dose distributions would lead to different outcomes. The cold part, which would greatly diminish the tumour control, is more detrimental than the hot spot.

In this paper we have established a general inverse planning framework in which the penalty at a voxel depends not only on the dose deviation from the desired value but also the dose-volume status of the involved organs. The technique circumvents the problems mentioned above and makes it possible to take advantage of the clinical outcome data. Our study shows that the incorporation of existing clinical knowledge can greatly facilitate the inverse planning process and allows us to obtain better IMRT plans. For the same target dose coverage, the critical structure sparing was substantially improved for both cases. Physically, we believe that the superior performance of the new formalism arises from the adequate modulation of the voxel dependent weighting induced by the dose-volume factor f (see equations (3) and (4)). In conventional dose-based objective function, $f \equiv 1$, and a tacit assumption that all points within a structure are equivalent has been made. The use of dose-volume factor f given by equation (3) enables us to weight different voxels according to the local doses. In this way, we can effectively 'boost' those target regions where the doses are low or penalize more those sensitive structure regions where the doses are high. The dose-volume induced voxel inequality is an important feature of the new inverse planning formalism and is the main driving-force in improving the dose distributions.

The use of clinical knowledge can also facilitate the determination of the structure specific importance factors. While the general influence of the importance factors on the solution is known, the specific response of the plan to a variation in the factors is not clear until the dose optimization is done, which necessitates a manual trial-and-error adjustment of the factors to achieve an acceptable trade-off. The underlying deficiency of the conventional approach is that the importance factors are purely heuristic and lack physical/clinical meanings. In this work we proposed a new scheme for modelling the trade-off and develop an algorithm to auto-determine the factors. The importance of an organ was written into a product of a generic and a dose-dependent factor. The latter was related to the corresponding TCP or NTCP. After the beam optimization, the dose-dependent factors were increased or decreased according to the values of TCPs/NTCPs. This procedure is similar to that reported by Xing *et al* (1999), where a DVH-based 'distance' was used for the assessment of the trade-off status after each optimization. In reality, other types of plan evaluation indices, such as mean/maximum/minimum doses, can also be employed for the purpose. We noted that, with the use of a new objective function, the final solution becomes much less sensitive to choice of $\{r_o^g\}$. This feature may have practical implications in simplifying the inverse planning process.

We should acknowledge that the new technique, just like any other dose response-based technique, may be limited by the scarcity and uncertainty of biological data and the limited predictive power of the TCP/NTCP models. It is important to note, however, that the TCP/NTCP is used as a relative ranking in our plan optimization algorithm instead of a clinical decision-making tool. Because of the phenomenological nature of the modelling, one may further modify the structure specific importance manually to achieve a certain clinical goal. The proposed technique can, at least, provide us with a good starting point for the fine-tuning. Our experience, along with the results shown in the above section, indicates that the technique proposed in this work is capable of generating clinically sensible plans and is much more efficient than the manual selection process.

Finally, we mention that the equivalent uniform dose (EUD)-based dose optimization (Niemierko 1997, Wu *et al* 2002, Thieke *et al* 2003, Lian and Xing 2004) can be cast into the realm of the above inverse planning framework and represents a special case of the general formalism described in this work. Indeed, assuming $EUD = (\frac{1}{N} \sum_i^N D_c^a(i))^{1/a}$, $a = 1/n$ and

setting k_τ and k_σ in equation (4) to zero, we can rewrite equation (4) into

$$F = \sum_{\tau=1}^{t_\tau} r_\tau \left[1 + \eta_1^t (\text{EUD}/\text{EUD}_{\tau,\text{ref}})^{a_\tau} + \eta_2^t (\text{EUD}/\text{EUD}_{\tau,\text{ref}})^{2a_\tau} + \dots \right] \\ + \sum_{\sigma=1}^{s_\sigma} r_\sigma \left[1 + \eta_1^s (\text{EUD}/\text{EUD}_{\sigma,\text{ref}})^{a_\sigma} + \eta_2^s (\text{EUD}/\text{EUD}_{\sigma,\text{ref}})^{2a_\sigma} + \dots \right], \quad (7)$$

which becomes a function of EUD. Different from the EUD-based model, the general hybrid objective function given in equation (6) treats the dose-volume effect at a more fundamental voxel level with the actual radiation dose considered, which is more flexible than the EUD defined at a structure level. Because of this, other clinical/dosimetric requirements can easily be integrated. Our study for the prostate case suggests that it is necessary to include the higher order contribution(s) if equation (3) or (7) is used to appropriately model a sensitive structure. Alternatively, a hybrid of dose-volume and dose-based objective function, as given by equation (6), can yield equally good plans. In practice, equation (6) is quite broad and seems to model the inverse planning system effectively. It may also find a natural application in functional image-guided IMRT, where the goal is generally to produce a spatially inhomogeneous dose distribution (Xing *et al* 2002). Finally, we note that the formalism does not involve the prescription of EUD, which could be problematic for practical implementation of an EUD-based model.

5. Conclusion

Inverse planning is an important step in IMRT and its performance crucially determines the quality of IMRT treatment plans. In this work, we provide a mechanism for incorporating clinical endpoint data into the inverse treatment planning process and established a clinically practicable inverse planning framework. We employed the effective volume in voxel domain to take the dose-volume effect of the involved organs into account. The new formalism gives important insight into the problem of therapeutic plan optimization. An algorithm for using computers to aid the determination of structure specific importance factors was also developed. A key step for accomplishing the auto-determination of the importance factors is the decomposition of the conventional importance factor into a generic importance and a dose-dependent component. Two case studies were presented to demonstrate the advantages of the proposed objective function. Comparison of the newly proposed approach with the conventional inverse planning technique indicated that the algorithm is capable of greatly improving the sensitive structure sparing with comparable target dose coverage and homogeneity.

Acknowledgments

Support from the National Cancer Institute (1R01 CA98523-01), the American Cancer Society (IGR-58-008-40), and Department of Defense (DAMD17-03-1-0019 and -0023) is gratefully acknowledged.

References

- Amols H I and Ling C C 2002 EUD but not QUD *Int. J. Radiat. Oncol. Biol. Phys.* **52** 1–2
- Brahme A 2001 Individualizing cancer treatment: biological optimization models in treatment planning and delivery *Int. J. Radiat. Oncol. Biol. Phys.* **49** 327–37

- Burman C, Kutcher G J, Emami B and Goitein M 1991 Fitting of normal tissue tolerance data to an analytic function *Int. J. Radiat. Oncol. Biol. Phys.* **21** 123–35
- Chen Y, Michalski D, Houser C and Galvin J M 2002 A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy *Phys. Med. Biol.* **47** 1647–58
- Cho P, Lee S, Marks R II, Oh S, Sutlief S and Phillips M 1998 Optimization of intensity modulated beams with volume constraints using two methods: cost function minimization and projections onto convex sets *Med. Phys.* **25** 435–43
- Deasy J, Niemierko A, Herbert D, Yan D, Jackson A, Ten Haken R, Langer M and Sapareto S 2002 Methodological issues in radiation dose–volume outcome analyses: summary of a joint AAPM/NIH workshop *Med. Phys.* **29** 2109–27
- Earl M A, Shepard D M, Naqvi S, Li X A and Yu C 2003 Inverse planning for intensity-modulated arc therapy using direct aperture optimization *Phys. Med. Biol.* **48** 1075–89
- Eisbruch A, Ten Haken R, Kim H, Marsh L and Ship J 1999 Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer *Int. J. Radiat. Oncol. Biol. Phys.* **45** 577–87
- Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrider J E, Shank B, Solin L J and Wesson M 1991 Tolerance of normal tissue to therapeutic irradiation *Int. J. Radiat. Oncol. Biol. Phys.* **21** 109–22
- Holmes T W, Mackie T R and Reckwerdt P 1995 An iterative filtered backprojection inverse planning algorithm for tomotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **32** 1215–25
- Hou Q, Wang J, Chen Y and Galvin J M 2003 An optimization algorithm for IMRT—the simulated dynamics with dose–volume constraints *Med. Phys.* **30** 61–8
- Hristov D, Stavrev P, Sham E and Fallone B 2002 On the implementation of dose–volume objectives in gradient algorithms for inverse planning *Med. Phys.* **29** 848–56
- Kallman P, Lind B K and Brahme A 1992 An algorithm for maximizing the probability of complication-free tumour control in radiation therapy *Phys. Med. Biol.* **37** 871–90
- Kutcher G J and Burman C 1989 Calculation of complication probability factors for non-uniform normal tissue irradiation: the effective volume method. *Int. J. Radiat. Oncol. Biol. Phys.* **16** 1623–30
- Lahanas M, Schreibmann E and Baltas D 2003 Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms *Phys. Med. Biol.* **48** 2843–71
- Langer M, Brown R, Kijewski P and Ha C 1993 The reliability of optimization under dose–volume limits *Int. J. Radiat. Oncol. Biol. Phys.* **26** 529–38
- Langer M, Morrill S and Lane R 1998 A test of the claim that plan rankings are determined by relative complication and TCP *Int. J. Radiat. Oncol. Biol. Phys.* **41** 451–7
- Lee E K, Fox T and Crocker I 2000 Optimization of radiosurgery treatment planning via mixed integer programming *Med. Phys.* **27** 995–1004
- Lian J and Xing L 2004 Incorporating model parameter uncertainty into inverse treatment planning *Med. Phys.* **31** 2711–20
- Llacer J, Solberg T and Promberger C 2001 Comparative behaviour of the dynamically penalized likelihood algorithm in inverse planning *Phys. Med. Biol.* **46** 2637–63
- Lyman J and Wolbarst A 1987 Optimization of radiation therapy. III: A method of assessing complication probabilities from dose–volume histograms *Int. J. Radiat. Oncol. Biol. Phys.* **13** 103–9
- Michalski D, Xiao Y, Censor Y and Galvin J 2004 The dose–volume constraint satisfaction problem for inverse planning with field segments *Phys. Med. Biol.* **49** 601–16
- Miften M M, Das S K, Su M and Marks L B 2004 Incorporation of functional imaging data in the evaluation of dose distributions using the generalized concept of EUD *Phys. Med. Biol.* **49** 1711–21
- Mohan R, Mageras G S, Baldwin B, Brewster L J, Kutcher G J, Leibel S, Burman C M, Ling C C and Fuks Z 1992 Clinically relevant optimization of 3D conformal treatments *Med. Phys.* **19** 933–44
- Mohan R, Wang X, Jackson A, Bortfeld T, Boyer A, Kutcher G, Leibel S A, Fuks Z and Ling C C 1994 The potential and limitations of the inverse radiotherapy technique *Radiother. Oncol.* **32** 232–48
- Niemierko A 1997 Reporting and analyzing dose distributions: a concept of equivalent uniform dose *Med. Phys.* **24** 103–10
- Schultheiss T E, Orton C G and Peck R A 1983 Models in radiotherapy: volume effects *Med. Phys.* **10** 410–5
- Shepard D, Earl M, Li X, Naqvi S and Yu C 2002 Direct aperture optimization: a turn-key solution for step-and-shoot delivery *Med. Phys.* **29** 1007–18
- Starkschall G, Pollack A and Stevens C 2001 Treatment planning using a dose–volume feasibility search algorithm *Int. J. Radiat. Oncol. Biol. Phys.* **49** 1419–27
- Thieke C, Bortfeld T, Niemierko A and Nill S 2003 From physical dose constraints to EUD constraints in inverse radiotherapy planning *Med. Phys.* **30** 2332–9

- Wang X H, Mohan R, Jackson A, Leibel S A, Fuks Z and Ling C C 1995 Optimization of intensity-modulated 3D conformal treatment plans based on biological indices *Radiother. Oncol.* **37** 140–52
- Webb S 2004 Direct aperture optimization for a variable aperture collimator for intensity-modulated radiation therapy *Phys. Med. Biol.* **49** N47–55
- Webb S and Nahum A E 1993 A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density *Phys. Med. Biol.* **38** 653–66
- Wu Q, Mohan R, Niemierko A and Schmidt-Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the EUD *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35
- Xing L, Cotrutz C, Hunjan S, Boyer A, Adalsteinsson E and Spielman D 2002 Inverse planning for functional image-guided IMRT *Phys. Med. Biol.* **47** 3567–78
- Xing L, Hamilton R J, Spelbring D, Pelizzari C, Chen G T and Boyer A L 1998 Fast iterative algorithms for 3D inverse treatment planning *Med. Phys.* **25** 1845–9
- Xing L, Li J G, Donaldson S, Le Q T and Boyer A 1999 Optimization of importance factors in inverse planning *Phys. Med. Biol.* **44** 2525–36
- Yan H, Yin F F, Guan H Q and Kim J 2003 AI-guided parameter optimization in inverse treatment planning *Phys. Med. Biol.* **48** 3565–80

**Dose-Volume Based Ranking of Incident Beam Direction and its
Utility in Facilitating IMRT Beam Placement**

Eduard Schreibmann, Ph.D. and Lei Xing^{a)}, Ph.D.

Department of Radiation Oncology, Stanford University School of Medicine,
Stanford, California 94305-5847, USA

^{a)}Author to whom correspondence should be addressed.

Stanford University School of Medicine
Department of Radiation Oncology
875 Blake Wilbur Drive
Stanford, CA 94305-5847

E-mail: lei@reyes.stanford.edu

Phone: (650) 498 7896

Fax: (650) 498 4015

Submitted to: Int. J. Radiat. Oncol. Biol. Phys.

Purpose: Beam orientation optimization in IMRT is computationally intensive and various single beam ranking techniques have been proposed to reduce the search space. Up to this point, none of the existing ranking techniques considers the clinically important dose-volume effects of the involved structures, which may lead to clinically irrelevant angular ranking. The purpose of this work is to develop a clinically sensible angular ranking model with incorporation of dose-volume effects and to show its utility for IMRT beam placement.

Materials and methods: The general consideration in constructing this angular ranking function is that a beamlet/beam is preferable if it can deliver a higher dose to the target without exceeding the tolerance of the sensitive structures located on the path of the beamlet/beam. In the previously proposed dose-based approach, the beamlets are treated independently and, to compute the maximally deliverable dose to the target volume, the intensity of each beamlet is pushed to its maximum intensity without considering the values of other beamlets. When volumetric structures are involved, the complication arises from the fact that there are numerous dose distributions corresponding to the same dose-volume tolerance. In this situation, the beamlets are not independent and an optimization algorithm is required to find the intensity profile that delivers the maximum target dose while satisfying the volumetric constraint(s). In this study, the behavior of a volumetric organ was modeled by using the equivalent uniform dose (EUD). A constrained sequential quadratic programming algorithm (CFSQP) was used to find the beam profile that delivers the maximum dose to the target volume without violating the EUD constraint(s). To assess the utility of the proposed technique, we planned a head and neck and abdominal case with and without the guidance of the angular ranking information. The qualities of the two types of IMRT plans were compared quantitatively.

Results: An effective angular ranking model with consideration of volumetric effect has been developed. It is shown that the previously reported dose-based angular ranking represents a special case of the general formalism proposed here. Application of the technique to a abdominal and a head and neck IMRT case indicated that the proposed technique is capable of producing clinically sensible angular ranking. In both cases, we found that the IMRT plans obtained under the guidance of EUD-based angular ranking were improved in comparison with that obtained using the conventional uniformly spaced beams.

Conclusions: The EUD-based function is a general approach for angular ranking and allows us to identify the potentially good and bad angles for clinically complicated cases. The ranking can be used either as a guidance to facilitate the manual beam placement or as prior information to speed up the computer search for the optimal beam configuration. Thus the proposed technique should have positive clinical impact in facilitating the IMRT planning process.

Key Words: IMRT, inverse planning, dose optimization, beam orientation.

INTRODUCTION

Several previous studies have indicated that the selection of beam directions plays an important role in IMRT planning (1-8). For a given patient, a practical challenge is how to obtain the optimal beam configuration within a clinical acceptable time. The influence of a candidate beam configuration to the final dose distribution is not known until an inverse planning with the configuration is done, which necessitates a trial-and-error process for the determination of a clinically sensible IMRT beam configuration (9-11). The use of computer optimization promises to automate the beam selection process (12-15). In practice, however, beam orientation optimization is computationally intensive because of the interplay between beam directions and beamlet intensities. Many researchers have attempted to use single beam ranking to reduce the search space (16-23). While such technique ignores the beam interplay and does not yield the final beam configuration, it affords useful information on which are potentially good/bad directions and is thus valuable to aid the beam placement (24-26). Furthermore, the information can be used as *a priori* knowledge to greatly improve the convergence behavior of the beam orientation optimization and to speed up the calculation process (21).

Single beam ranking functions can be divided into geometry based and dosimetry based models. The underlying difference between these models lies in what fundamental quantities are used to define the optimality. The former approach is essentially an extension of beam's eye view (BEV) volumetrics (24, 27, 28) and ranks a beam direction based on the fractional volume of sensitive structures falling into the incident beam (16, 22, 26). The dosimetry based technique, on the other hand, ranks a beam direction based on patient geometry as well as the *a priori* dose tolerance information of the involved sensitive structures (17, 18, 20). In line of the later approach, Pugachev and Xing proposed a heuristic

beam's eye view dosimetric (BEVD) model in which the ranking of an incident beam is gauged by the maximally deliverable dose to the planning target volume (PTV) without exceeding the tolerance(s) of the sensitive structure(s) located on the path of the beam (20). While the idea of their technique is quite general, practical implementation is hindered when the involved sensitive structure(s) is a parallel organ whose tolerance cannot be described by a single dose value but a dose-volume relation. In this situation, the method described in Ref. (20) is not directly applicable and the evaluation of the single beam ranking becomes very non-trivial.

The purpose of this work is to develop a clinically sensible angular ranking model with incorporation of dose-volume effect and to show its utility for IMRT beam placement. Instead of working in the dose domain, we establish a more adequate formalism based on the concept of equivalent uniform dose (EUD) (29, 30). As is well known, the dose-volume effect of a structure is well described (31-33) by using the EUD, which is phenomenologically defined as:

$$EUD = \left(\frac{1}{N} \sum_i D_i^a \right)^{\frac{1}{a}}, \quad (1)$$

where N is the number of voxels in the structure, a is the tumor or normal tissue-specific parameter that describes the dose-volume effect. When $a = \infty$, the EUD is equal to the maximal dose, and when $a = -\infty$, the EUD is equal to the minimum dose. Tumors generally have large negative values of a , whereas serial critical structures (e.g. spinal cord) have large positive values and parallel critical structures that exhibit a large volumetric effect (e.g. liver, parotids, and lungs) have small positive values. A virtue of the EUD based approach is that the volumetric status of an organ can be characterized by a single parameter instead of two (dose and volume), making the incorporation of dose-volume effect in the single beam

ranking more tractable. A few of recent work (34-36) have attempted to relate EUD-based formalism to dose-volume constraints, dose-based inverse planning, and multiobjective optimization and significant insights have been obtained in this regard.

METHOD

Background

The figure of merit of a beam direction is generally measured by how much dose can be delivered to the target and is calculated using the *a priori* dosimetric and geometric information of the given patient. For computational purpose, a beam direction is divided into a grid of beamlets. For a serial organ, each beamlet crossing the target is an independent element and is assigned to the maximum intensity (20). After a forward dose calculation using the maximum beam intensity profile, the score of the given beam direction (indexed by b) is obtained according to

$$S_b = \frac{1}{N_T} \sum_{n \in \text{Target}} \left(\frac{d_{nb}}{D_T^P} \right)^2, \quad (2)$$

where d_{nb} is the dose delivered to the voxel n by the beam from the direction indexed by b , N_T the number of voxels in the target, and D_T^P the target prescription.

When the tolerance of the sensitive structure is more accurately assessed by volumetric information, a modification of the above procedure is needed to rank the beam direction. The main computational difficulty here is that the beamlets cannot be dealt independently as there are usually multiple dose distributions for a given dose-volume tolerance. A practical and efficient approach is to use the EUD model described earlier to solve the problem.

EUD as a general means for characterizing the dose-volume tolerance

The dose volume tolerance of a sensitive structure is more conveniently described by a single variable, namely, the EUD tolerance. The value of EUD tolerance (and the value of a in the EUD formula (1)) for a given organ can be obtained from the literature (37). For a given case, the figure of merit of a beam direction should now be measured by what the beam could achieve dosimetrically without exceeding the EUD tolerance(s) of the organ(s) located on the path of the beam. While the score can still be evaluated by using Eq. (1) based on the integral dose to the target, the method of determining the maximum intensity profile for a given beam direction must be modified. In this situation, the beamlet intensities of the maximum intensity profile can no longer be determined individually because of the degenerate nature of a dose-volume tolerance (*i.e.*, there are more than one dose distributions for a given dose-volume tolerance). Generally, there are multiple intensity profiles, each corresponding to a possible dose distribution of the given EUD tolerance. A computer algorithm is necessary to find the one that delivers the maximal integral dose to the target volume. For this purpose, we mathematically formulated the problem into a constrained optimization, in which the objective is to maximize S_b defined in Eq. (2) with the EUDs of the sensitive structures below their tolerance values.

Calculation of beam ranking

The constrained optimization problem was solved using a Lagrange multiplier method (38). For the system that we are dealing here, the Lagrange function takes the form of

$$L(\mathbf{x}_b, \lambda) = S_b(\mathbf{x}_b) + \sum_i \lambda_i [EUD_{tol}^i(\mathbf{x}_b) - EUD^i(\mathbf{x}_b)], \quad (3)$$

where \mathbf{x}_b is the system variable vector (*i.e.*, the beamlet intensities for a given beam), i is the structure index, b is the beam index, $EUD_{tol}^i(\mathbf{x}_b)$ and $EUD^i(\mathbf{x}_b)$ are the EUD tolerance and computed EUD, respectively, and λ_b^i is the Lagrange multiplier. A constrained sequential quadratic programming algorithm (CFSQP) (39) with the Armijo type arc search mode (40) was employed to optimize the system described above. Variables to be optimized included the intensities of the beamlets passing through the PTV. A major advantage of the algorithm is that it is capable of dealing with nonlinear inequality constraints (40). The calculation starts with an initial intensity profile, in which each beamlet is assigned with a small but random value, and then iteratively maximizes the angular ranking function while satisfying the constraints. Briefly, the k -th iteration can be described by the following steps:

1. Calculate the inverse Hessian matrix, \mathbf{H}_k , from the gradient of $L(\mathbf{x}, \lambda)$;
2. Determine the descent direction $\mathbf{p}_k = -\mathbf{H}_k \nabla f(\mathbf{x}_k)$;
3. Line search with a step size $\alpha_k = \arg \min_{\alpha > 0} f(\mathbf{x}_k + \alpha \mathbf{p}_k)$, where α is the step size defined in the CFSQP software package;
4. Update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$;
5. Check stopping criteria. If not satisfied, repeat from 1;

CFSQP calculation stops either when a maximum number of 200 iterations is reached or when the difference in the values of the ranking function between two successive iterations becomes less than 10^{-6} . All calculations are implemented in an in-house treatment planning platform (6).

It is important to emphasize that the goal of the above calculation is not to find the optimal IMRT solution. Instead, we are searching for the beam profiles that deliver the highest achievable dose in the target without violating the dose-volume or EUD tolerances of

the involved sensitive structures. In the final solution any increase in the beamlet weights would lead to a dose exceeding the volumetric tolerance of a sensitive structure.

Optimization of IMRT beam profiles

The flow chart of plan optimization process is shown in figure 1. We used a hybrid of a gradient-based algorithm (LBFGS) and a genetic like algorithm (NSGA-IIc)(41) to obtain the optimal IMRT plan after the beam configuration is selected. The LBFGS is specially designed for dealing with systems with a large number of variables (42). A least squares cost function was used as the objective for each structure, with the prescription dose set to 0 for sensitive structures in order to continuously lower the doses to the sensitive structures if there is room for improvement. The details of the LBFGS algorithm and our experience in implementing the algorithm were reported in a previous work referenced in the manuscript (43). The EUDs of the involved structures were employed to generate the solution pool through the combinational use of a genetic like algorithm (NSGA-IIc)(41) (see figure 1 for the flow chart for detail). At each iteration, the EUDs of the new solution are compared with the plan within the solution pool and ranked according the rules stated in the flow chart. A solution that violates less the EUD constraints have a higher probability to survive in the next generation and thus a larger chance to be selected as the final solution. In this sense, the EUDs here act as “soft constraints” in conventional optimization or the decision-function in the two-step optimization proposed by Xing et al (44). This setup has the advantage that the EUD constraints do not directly interfere with the objective function of the system. This model of implementing the constraints was found to be efficient in a previous study (6).

Case studies

The performance of the above algorithm and the advantage of the proposed angular ranking technique are illustrated using two clinical IMRT cases, a abdominal tumor with the kidney, spinal cord and liver delineated as sensitive structures and a head and neck case adjacent to the parotid and spinal cord. For each case, we computed the coplanar angular score function (from 0° to 360°) in an increment of 5° . The photon beam energy was 15 MV for both cases.

To understand the behavior of the technique in modeling different degree of volumetric effect, we computed the score function for a series of values of parameter a in each case. As a increases to a large positive value, the organ behaves as a serial structure and the score function obtained by using the proposed technique should converge to the results obtained using the method outlined in Ref. (20).

To illustrate the advantage of the technique, we generated a five-field IMRT treatment plan with the beam configuration obtained under the guidance of the computed angular score function for each patient. The plans were then compared with that obtained with five uniformly spaced beams (the beam directions are 0° , 72° , 144° , 216° , and 288°). In practice, 5~7 beams are often used for IMRT treatment because a good compromise can be achieved between delivery efficiency and the treatment quality. After the beam configuration was determined, the beamlets were optimized using a multi-objective approach described above (see Fig. 1) to yield the final IMRT plan. The optimization result is a database of plans representing the best achievable tradeoff between optimization objectives. The DVHs of the target and sensitive structures were used for the plan comparison.

In order to better assess the performance of the proposed technique, two additional experiments were carried out for the head and neck case. First, we intentionally selected the

set of beams with the worst (corresponding to the dips of the beam orientation ranking function) and the best (corresponding to the peaks of the beam orientation ranking function, as chosen in the above) individual performances. The resulting IMRT plans were then compared quantitatively. This study highlights the importance of beam selection and the possible influence of individually selected beams. For the abdominal patient, we also investigated the differences in the final IMRT treatment plans obtained using the new method with EUD-based angular score function and the previous non-EUD based score function.

RESULTS

Abdominal case

The angular score functions corresponding to different values of the α parameter are presented in figure 2a along with the score obtained using the method reported in Ref. (20) for serial organ(s). As can be intuitively anticipated, the EUD-based angular score function approaches to the latter curve (denoted by the circles) as α increases. The values of EUD tolerances for the involved organs were taken from the literature (Table 1) (45). The score function corresponding to the α parameters extracted from Ref. (45) is also presented in figure 2b, where a polar graph has been overlayed on the patient's geometry to better interpret the curve. The influence of the sensitive structures is clearly reflected in the score function. The lowest scores occur at the angles close to 50° and 320° , from which directions the beam pass through the kidneys and the liver. The score function peaks at directions where a large portion of PTV can be irradiated with minimal intersection of sensitive structures.

A general way for beam placement after the BEVD score is obtained is to start with an equispaced beam configuration with one of the beams placed at the angle with the highest

BEVD score value. The angular score described above is obtained under the assumption of a single incident beam and thus reflects only one facet of the beam configuration selection problem. The angles of the rest of beams are adjusted in such a way that the final beam configuration balances the requirement of the single beam BEVD score (which favors placing beams at the peak positions of the BEVD score function) and the interplay between the beams (which prefers a beam configuration with the beams away from each other). In other words, to obtain the final beam configuration, it is required to consider the interplay between the incident beams. The merit of the proposed technique is that it is capable of identifying potentially "good" and "bad" directions even when the sensitive structures are volumetric in their responses to radiation, and thus provides a useful guidance for IMRT beam placement. In particular, it permits us to avoid placing beams in the "bad" beam directions (the dips of the score function). One would otherwise have to choose a beam configuration entirely based on empirical "guessing". Based on the above consideration, we placed the five incident beams at (35°, 95°, 190°, 280°, 335°), as indicated by the arrows in figures 2a and 2b. The DVHs of the corresponding IMRT plan are plotted in figure 2c as dot-dashed lines. The DVHs of a conventional IMRT plan with five equispaced beams are also plotted in figure 2c as solid curves.

The computed EUD values of the involved structures for IMRT plans with equispaced beams, non-EUD-based BEVD-selected beams and newly proposed EUD-based BEVD-selected beams are also listed in Table1. We observed that the IMRT plan with the beam configuration selected under the guidance of the EUD-based scoring function represents the best treatment when judged using the EUDs of the involved structures. This is not surprising since the EUDs of various organs were fully considered during the beam configuration selection. The IMRT plan with beams derived from the non-EUD based score function (the

five beam directions derived from this score function were 20°, 120°, 175°, 260°, and 350°) was inferior to the plan mentioned above but better than that of equispaced beam configuration. As seen from the table and figure 2c, the liver and both kidneys sparing are improved for the same PTV coverage. For example, the fractional volume of the right kidney receiving a dose of 15 Gy is reduced from 37% and 23% and the reductions in lower doses are even significant. In the IMRT plan with BEVD-selected beam configuration, however, the maximum dose to the spinal cord is increased slightly. This is because a relative high tolerance EUD was assigned to the cord, which therefore imposed less constraint on the angular ranking when the beam passes through it. For this case, the most dose-limiting organ is the left kidney when one attempts to escalate the target dose. Hence the slight increase in the cord dose has little clinical impact. In a sense, the increase in the cord dose is an indication that the newly introduced ranking function can more effectively balance the requirements of the involved structures.

Convergence behavior of the CFSQP algorithm is demonstrated by plotting the angular score as a function of iteration step (figure 3a) for the 255° direction. The EUDs of the sensitive structures at each iteration step are shown in figure 3b. With the chosen initial beamlet intensities (small but random values), the angular score is progressively increased while constraints are progressively saturated, limited by the tolerances of the sensitive structures. Constraints of the sensitive structures that are not on the path of the beam remain to be constant throughout the iterative calculation. In this particular situation, the right kidney is the most dose-limiting organ. Of course, being dose-limiting for one of the beams does not mean that the organ is dose-limiting when all the beams are considered. Indeed, as seen from figure 2b and table 1, it turns out that the left kidney is most dose-limiting for the chosen beam configuration and receives the highest dose or EUD. Generally speaking, this is not a

symmetric system from the dosimetric point of view, even though the two kidneys have the same EUD tolerance. The asymmetric behavior is caused primarily by the presence of the liver near the right kidney. When the dose to the liver is limited to its EUD tolerance, a by-consequence is that the dose to the right kidney is also lowered. On the other hand, the left kidney is the sole dose limiting organ on the left side and a much higher dose is delivered to it in order to deliver the prescribed dose to the target volume. The asymmetry here can also be seen from the angular ranking function (figure 2a).

Head and neck case

Dependence of EUD score on the a parameter is presented in figure 4a. As a increases, the angular score curve approaches to the curve computed using the method outlined in Ref. (20) (denoted by the open circles). This calculation provides a useful check of the new algorithm. It is interesting to note that the change in the peak positions of the angular function can be as large as 20° when the sensitive structures are changed from serial (corresponding to a high a value) to parallel (corresponding to a low a value). The change of the score function in amplitude is also striking (from ~ 0.2 to ~ 0.8 at 80° and 280°). As the structure becomes more volumetric (e.g., when $a=0.1$), the score function becomes less "spiky".

When the published values of EUD tolerances (45) are used (table 2), the lateral directions passing through parotid are disadvantageous, while the anterior-posterior directions passing through the cord have slightly higher rankings. The polar score function overlayed on the patient's geometry is presented in figure 4b. It is clearly seen that low scores occur at $\sim 90^\circ$ and $\sim 270^\circ$, from which directions the beams pass through the parotids. Moderate scores appear at 0° and 180° , from which directions the beams pass through the cord. In this case, as

can be intuitively imagined, high angular rankings appear in the oblique directions when the beam is less intercepted by the sensitive structures.

Under the guidance of the ranking function depicted in figure 4a, we selected (45° , 125° , 190° , 245° , 310°) for the five-beam IMRT treatment. The selected directions are indicated in figures 4a and 4b by arrows. The DVHs of the involved structures are shown in figure 4c (dot-dashed lines) along with the results obtained by using five uniformly spaced beams. As a result of BEVD guidance for beam placement, the IMRT plan is improved significantly. Together with a dose improvement in dose coverage of the PTV, all sensitive structures are better spared. The improvement in brainstem sparing is most distinct as compared with the conventional plan. The resulting EUD of the brainstem was reduced by $\sim 65\%$ (see table 2). The reduction in the parotid EUD was found to be $\sim 16\%$. The fractional volume of the spinal cord receiving a dose of 14 Gy is reduced from 32% to 22%. Clinically, the improvement in the dose homogeneity within the PTV may have some practical implications in enhancing the tumor control probability. In addition, better sparing of the involved sensitive structures yields room for possible boost treatment of the patient, or, in the case of recurrence, for the re-treatment. The convergence behavior of the optimization calculation for the 180° direction is presented in figure 5. Once again, it is seen that the calculation converges rapidly. The constraints imposed by the parotid glands are satisfied from the beginning of the iterative calculation because the beam does not pass them.

In figure 6 we compare the DVHs for IMRT plans obtained with five beams placed at the best and worst individual performance directions as revealed by the BEVD score function. The worst beam directions were identified to be (25° , 80° , 170° , 280° , and 340°). As seen from figure 6, the PTV dose was deteriorated uniformly when the worst beam configuration are used for IMRT planning. Simultaneously, the doses to all the sensitive structures are increased

as compared to that with best individual performance beams. The computed EUDs listed in table 2 also support the above observation. This comparison study underscores the importance of beam orientation selection and provides additional evidence that the BEVD is a useful tool to facilitate IMRT planning.

DISCUSSION

Increasing the effectiveness of radiation therapy depends on improving our understanding of dose-volume factors affecting tolerance and local control and developing techniques for incorporating the volumetric data into IMRT treatment plan optimization. To rank an incident beam direction in the presence of volumetric organs, a conceptually simple approach is to sample the tolerance dose distributions of the structure(s) according to the dose-volume tolerance requirement and to compute the score for each sampling (46, 47). With multiple sampling, we can obtain a "band" of scores rather than a single score curve. The best achievable target dose corresponds to the top of the band, which can be used for guiding the beam placement. While conceptually simple, sampling requires a large amount of computation and is difficult to implement. In this work we proposed a more practical and efficient EUD model-based approach (29, 30) to deal with the dose-volume effect. A major feature of the EUD formalism is that it allows us to use a single parameter to capture the dosimetric or volumetric characteristics of an organ. Several studies (34, 36, 48-50) have showed the utility of the formalism in plan optimization.

One of the technical hurdles in calculating the EUD-based angular score ranking is that, for a given set of EUD tolerances, there are multiple intensity profiles. The one that yields the highest score function (2) was obtained using a constrained optimization algorithm. We found that the CFSQP implemented in this work was capable of quickly converging to a

stable solution for the cases studied. As described in previous section, the IMRT plans obtained with beam configurations chosen based on the EUD-based score compares favorably with that obtained using other approaches, including equispaced beams and beams obtained under the guidance of a non-EUD based score function. We note that, because the non-EUD based score function represents a special case of the general EUD-based function, there may be cases where the EUD-based score yield the same IMRT plan as the non-EUD-based score (when all the organs involved are serial). However, it is important to keep in mind that the EUD-based function is more general approach and is more adequate when volumetric organs are involved.

Generally speaking, the optimal beam configuration needs to balance the single beam angular ranking and the beam interplay. The former favors placing beams at the peak positions of the BEVD score function, whereas the later prefers a beam configuration with the beams away from each other. While the two factors are not always conflicting each other, the beams other than the one placed at the highest peak often need to be moved away from some BEVD peaks to separate the beams. We admit that there exists a certain uncertainty in the placements of these beams. However, the uncertainty in this process is far less when compared with the situation where no BEVD information is available. In other words, the introduction of angular score function significantly reduces the size of the beam configuration search space and enables us to quickly obtain a set of clinically sensible beams. The results of this study demonstrated the effectiveness of a newly developed algorithm that incorporates the volumetric information of the involved organs. Further investigation into this technique should be performed to investigate other types of clinical cases and to develop full beam orientation optimization method to take advantage of the *a priori* knowledge about the angular search space as derived from the BEVD score evaluation (21).

CONCLUSION

An important issue in IMRT beam orientation ranking is how to incorporate the existing clinical endpoint data into the calculation. In this work we have described an effective method for incorporating the dose-volume effects into the evaluation of single beam ranking. In the proposed technique, the volumetric effect of an organ is characterized by the EUD and the ranking of a beam direction is assessed by the dose delivered to the PTV without exceeding the EUD tolerances or dose-volume limits of the sensitive structures. Because of the degenerate nature of the volumetric tolerance, an optimization algorithm is needed to find the maximum beam intensity profile from the pool of candidate beams that produce the same EUD. The EUD-based angular score allows a fast evaluation of all accessible beam angles based on the *a priori* volumetric and geometric information of the patient. Dosimetric or geometrical scoring schemes proposed previously by different investigators represent special cases of the general formalism here. Application of the reported technique to two clinical IMRT cases indicates that the method is capable of identifying potentially good and bad directions and useful to guide the beam placement process. The information can also be used as prior knowledge to facilitate the full beam orientation optimization process (21). Considering the relative simplicity of the model and efficiency of the calculation, the proposed technique should be a useful addition to the armamentarium of IMRT planning.

ACKNOWLEDGEMENTS

We would like to thank the useful discussion with Drs. A. Pugachev, Y. Yang, and C. Cotrutz. This work is supported in part by a research grant from the U.S. Department of Defense (DAMD17-03-1-0023) and National Cancer Institute (5 R01 CA98523-01).

REFERENCES

1. Stein J, Mohan R, Wang XH, Bortfeld T, Wu Q, Preiser K, Ling CC, Schlegel W. Number and orientations of beams in intensity-modulated radiation treatments. *Med Phys* 1997;24:149-60.
2. Pugachev AB, Boyer AL, Xing L. Beam orientation optimization in intensity-modulated radiation treatment planning. *Med Phys* 2000;27:1238-45.
3. Das S, Cullip T, Tracton G, Chang S, Marks L, Anscher M, Rosenman J. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. *Int J Radiat Oncol Biol Phys* 2003;55:215-24.
4. Djajaputra D, Wu Q, Wu Y, Mohan R. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Phys Med Biol* 2003;48:3191-3212.
5. Meedt G, Alber M, Nüsslin F. Non-coplanar beam direction optimization for intensity-modulated radiotherapy. *Phys Med Biol* 2003;48:2999-3019.
6. Schreibmann E, Lahanas M, Xing L, Baltas D. Multiobjective evolutionary optimization of the number of beams, their orientations and weights for intensity-modulated radiation therapy. *Phys Med Biol* 2004;49:747-770.

7. Gaede S, Wong E, Rasmussen H. An algorithm for systematic selection of beam directions for IMRT. *Med Phys* 2004;31:376-388.
8. Li Y, Yao J, Yao D. Automatic beam angle selection in IMRT planning using genetic algorithm. *Phys Med Biol* 2004;49.
9. Nutting CM, Rowbottom CG, Cosgrove VP, Henk JM, Dearnaley DP, Robinson MH, Conway J, Webb S. Optimisation of radiotherapy for carcinoma of the parotid gland: a comparison of conventional, three-dimensional conformal, and intensity-modulated techniques. *Radiother Oncol* 2001;60:163-72.
10. Earl MA, Shepard DM, Naqvi S, Li XA, Yu CX. Inverse planning for intensity-modulated arc therapy using direct aperture optimization. *Phys Med Biol* 2003;48:1057-1089.
11. Stavrev P, Hristov D, Warkentin B, Sham E, Stavreva N, Fallone BG. Inverse treatment planning by physically constrained minimization of a biological objective function. *Med Phys* 2003;30:2948-2958.
12. Wang C, Dai J, Hu Y. Optimization of beam orientations and beam weights for conformal radiotherapy using mixed integer programming. *Phys Med Biol* 2003;48:4065-4076.
13. Gokhale P, Hussein EM, Kulkarni N. Determination of beam orientation in radiotherapy planning. *Med Phys* 1994;21:393-400.
14. Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Med Phys Biol* 1993;38:291-304.

15. Hosseini-Ashrafi ME, H. B, Yahaqi E. Pre-optimization of radiotherapy treatment planning: an artificial neural network classification aided technique. *Phys Med Biol* 1999;44:1513-28.
16. Haas OC, Burnham KJ, Mills JA. Optimization of beam orientation in radiotherapy using planar geometry. *Phys Med Biol* 1998;43:2179-93.
17. Rowbottom CG, Webb S, Oldham M. Improvements in prostate radiotherapy from the customization of beam directions. *Med Phys* 1998;25:1171-9.
18. Rowbottom CG, Oldham M, Webb S. Constrained customization of non-coplanar beam orientations in radiotherapy of brain tumors. *Phys Med Biol* 1999;44:383-99.
19. Woudstra E, Storchi PRM. Constrained treatment planning using sequential beam selection. *Phys Med Biol* 2000;45:305-328.
20. Pugachev A, Xing L. Pseudo beam'-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 2001;51:1361-1370.
21. Pugachev A, Xing L. Incorporating prior knowledge into beam orientation optimization. *Int J Radiat Oncol Biol Phys* 2002;54:1565-1574.
22. Schreibmann E, Lahanas M, Uricchio R, Theodorou K, Kappas C, Zamboglou N, Baltas D. A geometry based optimisation algorithm for conformal external beam orientation. *Phys Med Biol* 2003;48:1825-1841.
23. Lee KE, Fox T, Crocke I. Integer Programming Applied to Intensity-Modulated Radiation Therapy Treatment Planning. *Ann Op Res* 2003;119:165-181.

24. Chen GT, Spelbring DR, Pelizzari CA, Balter JM, Myriantopoulos LC, Vijayakumar S, Halpern H. The use of beam's eye view volumetrics in the selection of non-coplanar radiation portals. *Int J Radiat Oncol Biol Phys* 1992;23:153-63.
25. Das SK, Marks LB. Selection of coplanar or noncoplanar beams using three-dimensional optimization based on maximum beam separation and minimized nontarget irradiation. *Int J Radiat Oncol Biol Phys* 1997;38:643-55.
26. Cho BJ, Roa WH, Robinson D, Murry B. The development of target-eye-view maps for selection of coplanar or noncoplanar beams in conformal radiotherapy treatment planning. *Med Phys* 1999;26:2367-2372.
27. Myriantopoulos LC, Chen GT, Vijayakumar S, Halpern HJ, Spelbring DR, Pelizzari CA. Beam's eye view volumetrics: an aid in rapid treatment plan development and evaluation. *Int J Radiat Oncol Biol Phys* 1992;23:367-375.
28. McShan DL, Kessler ML, Fraass BA. Advanced interactive planning techniques for conformal therapy: high level beam descriptions and volumetric mapping techniques. *Int J Radiat Oncol Biol Phys* 1995;33:1061-72.
29. Niemierko A. A generalized concept of equivalent uniform dose (EUD). *Med Phys* 1999;26:1100.
30. Niemierko A. Reporting, and analyzing dose distributions: A concept of equivalent uniform dose. *Med Phys* 1997;1:103-110.
31. Kavanagh BD, Timmerman RD, Benedict SH, Wu Q, Schefter TE, Stuhr K, McCourt S, Newman F, Cardinale RM, Gaspar LF. How should we describe the radiobiologic effect

of extracranial stereotactic radiosurgery: equivalent uniform dose or tumor control probability? Med Phys 2003;30:321-4.

32. Ebert MA. Viability of the EUD and TCP concepts as reliable dose indicators. Phys Med Biol 2000;45:441-57.
33. Wang JZ, Li XA. Evaluation of external beam radiotherapy and brachytherapy for localized prostate cancer using equivalent uniform dose. Med Phys 2003;30:34-40.
34. Yang Y, Xing L. Clinical knowledge-based inverse treatment planning. Phys Med Biol 2004;49:5101-5117.
35. Romeijn HE, Dempsey JF, Li JG. A unifying framework for multi-criteria fluence map optimization models. Phys Med Biol 2004;49:1991-2013.
36. Thieke C, Bortfeld T, Niemenko A, Nill S. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. Med Phys 2003;9:2332-2339.
37. Emami B, Lyman J, Brown A, Coia L, Goitein M, Munzenrider JE, Shank B, Solin LJ, Wesson M. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys 1991;21:109-22.
38. Crooks SM, Xing L. Application of constrained least-squares techniques to IMRT treatment planning. Int J Radiat Oncol Biol Phys 2002;4:1217-1224.
39. Lawrence CT, Tits AL. Computationally Efficient Feasible Sequential Quadratic Programming Algorithm. SIAM J Optimization 2001;11:1092-1118.
40. Panier E, Tits E. On Combining Feasibility, Descent and Superlinear Convergence In Inequality Constrained Optimization. Math Prog 1993;59:261-276.

41. Deb K, Goel T. Controlled elitist non-dominated sorting genetic algorithms for better convergence. Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science 2001;1993: 67-81.
42. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. Math Program 1989;45:503-528.
43. Lahanas M, Schreibmann E, Baltas D. Constrained Free Gradient-Based Optimization Algorithms for Multiobjective Inverse Planning in Intensity Modulated Radiotherapy. Phys Med Biol 2003;48:2843-2871.
44. Xing L, Li JG, Donaldson S, Le QT, Boyer AL. Optimization of importance factors in inverse planning. Phys Med Biol 1999;44:2525-36.
45. Song W, Dunscombe P. EUD-based margin selection in the presence of set-up uncertainties. Med Phys 2004;31:849-59.
46. Langer M, Brown R, Urie M, Leong J, Stracher M, Shapiro J. Large scale optimization of beam weights under dose-volume restrictions. Int J Radiat Oncol Biol Phys 1990;18:887-93.
47. Starkschall G, Pollack A, Stevens CW. Treatment planning using a dose-volume feasibility search algorithm. Int J Radiat Oncol Biol Phys 2001;49:1419-1427.
48. Wu Q, Mohan R, Niemierko A, Schmidt-Ullrich R. Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose. Int J Radiat Oncol Biol Phys 2001;52:224-235.

49. Chapet O, Thomas E, Kessler ML, Fraass BA, Haken RKT. Esophagus sparing with IMRT in lung tumor irradiation: An EUD-based optimization technique. Int J Radiat Oncol Biol Phys 2005;In Press.
50. Lian J, Xing L. Incorporating model parameter uncertainty into inverse treatment planning. Med Phys 2004;31:2711-20.

FIGURES CAPTIONS

Figure 1. Flow chart of the overall plan optimization process.

Figure 2. (a) Angular ranking function for a series of α parameters. The selected five directions for IMRT planning are labeled by arrows. The curve depicted by the open circles represents the result obtained using the approach described in Ref. (20). (b) Angular score obtained with published EUD tolerances and model parameters superimposed on the patient's geometry. Angles selected for IMRT planning are shown by arrows. (c) DVHs corresponding to the IMRT plans obtained with (dot-dashed curves) and without (solid) the guidance of the angular ranking function.

Figure 3. Convergence behavior of the CFSQP algorithm for the 255° beam direction. Presented are evolutions of the angular ranking (a) and the sensitive structure constraints (b) as function of iteration step. Only the right kidney and the liver influence algorithm's convergence since other structures are not on the path of the beam.

Figure 4. (a) Angular ranking function for a series of values of EUD α parameters. The selected five directions for IMRT planning are labeled by arrows. The curve depicted by the open circles represents the result obtained using the approach described in Ref. (20). (b) Angular score obtained with published EUD tolerances and model parameters superimposed on the patient's geometry. Angles selected for IMRT planning are shown by arrows. (c)

DVHs corresponding to the IMRT plans obtained with (dot-dashed curves) and without (solid) the guidance of the angular ranking function.

Figure 5. Convergence behavior of the CFSQP algorithm for the 180° beam direction. Presented are evolutions of the angular ranking (a) and the sensitive structure constraints (b) as function of iteration step. The cord and brainstem limit the dose deliverable to the PTV in this beam direction.

Figure 6. Comparison of DVHs of IMRT plans with the best and worst individual performance directions as identified by the angular ranking function.

FIGURE 1

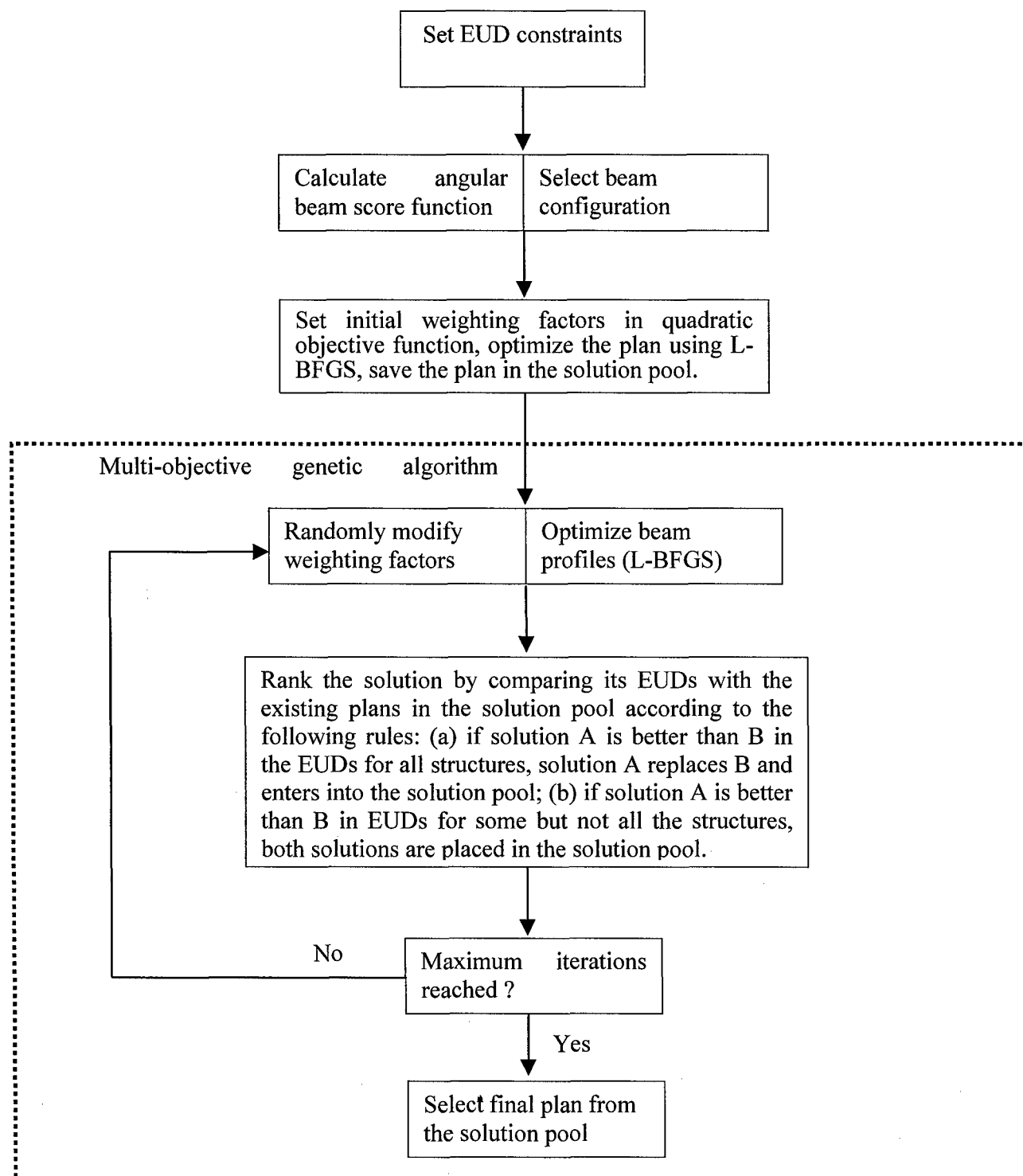
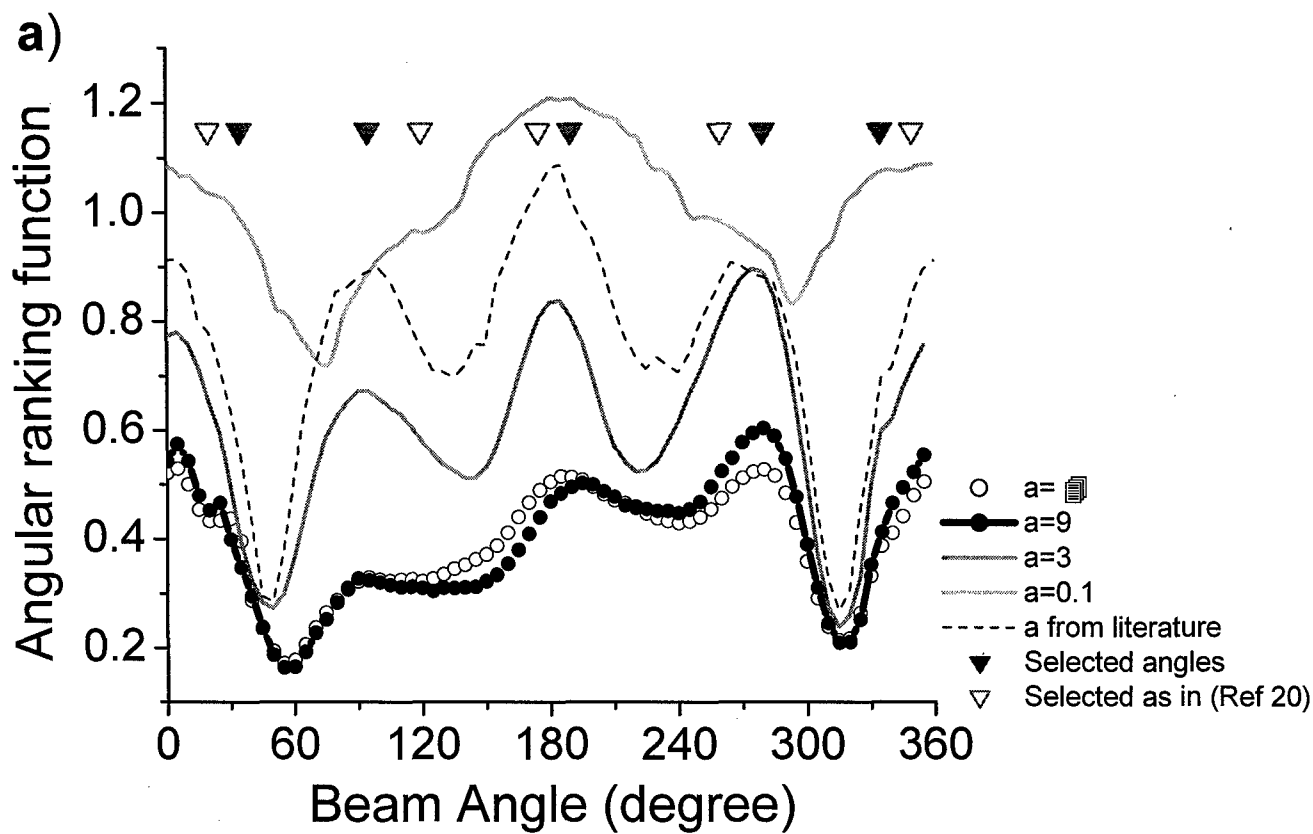
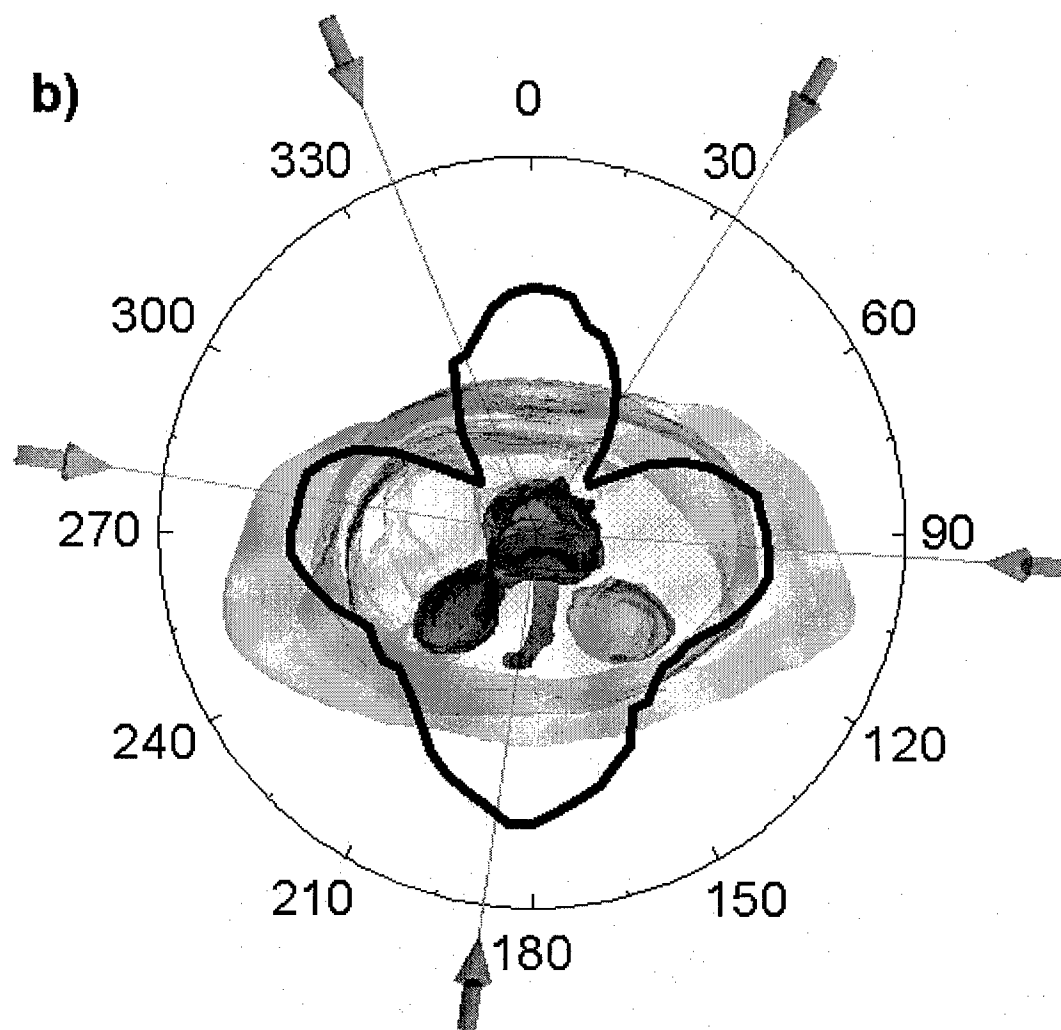


FIGURE 2





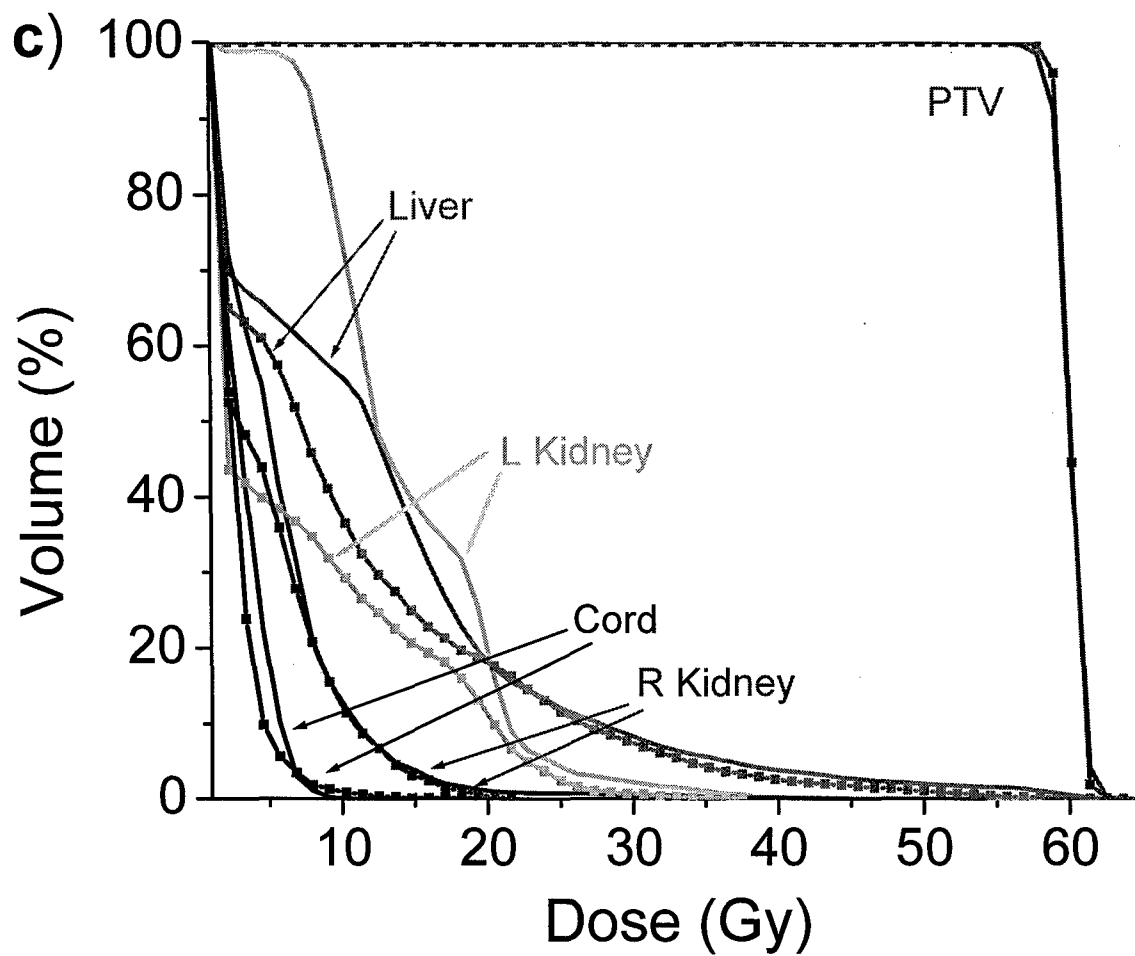


FIGURE 3

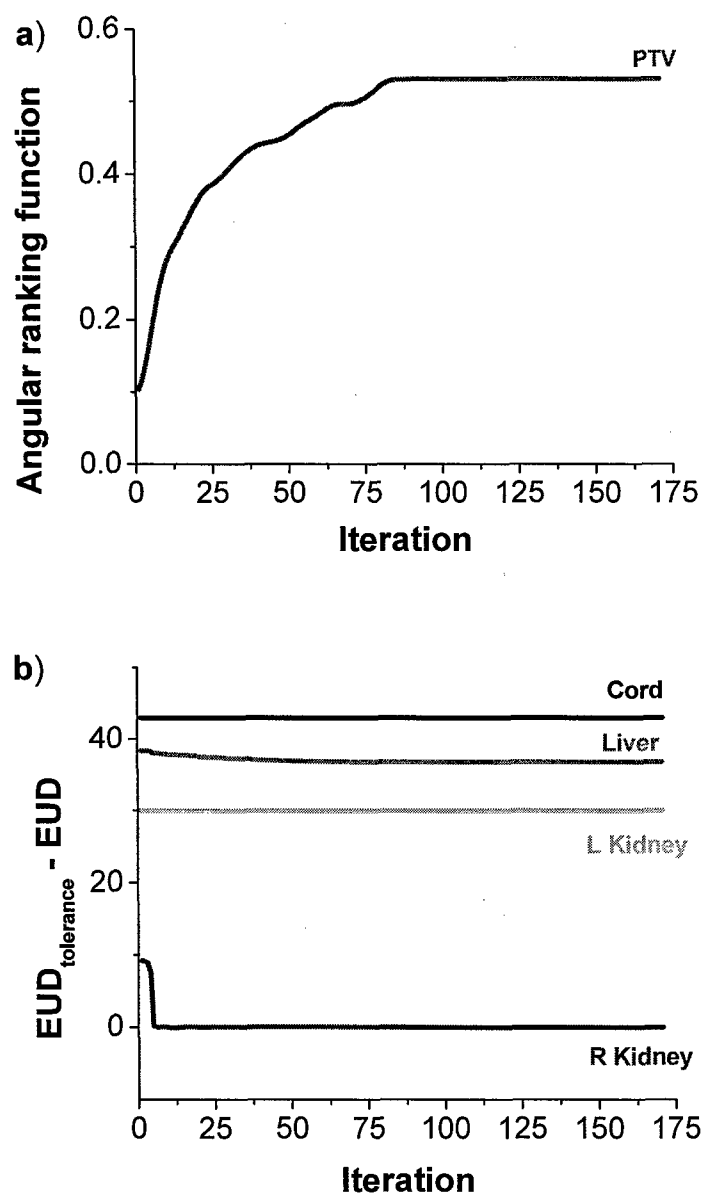
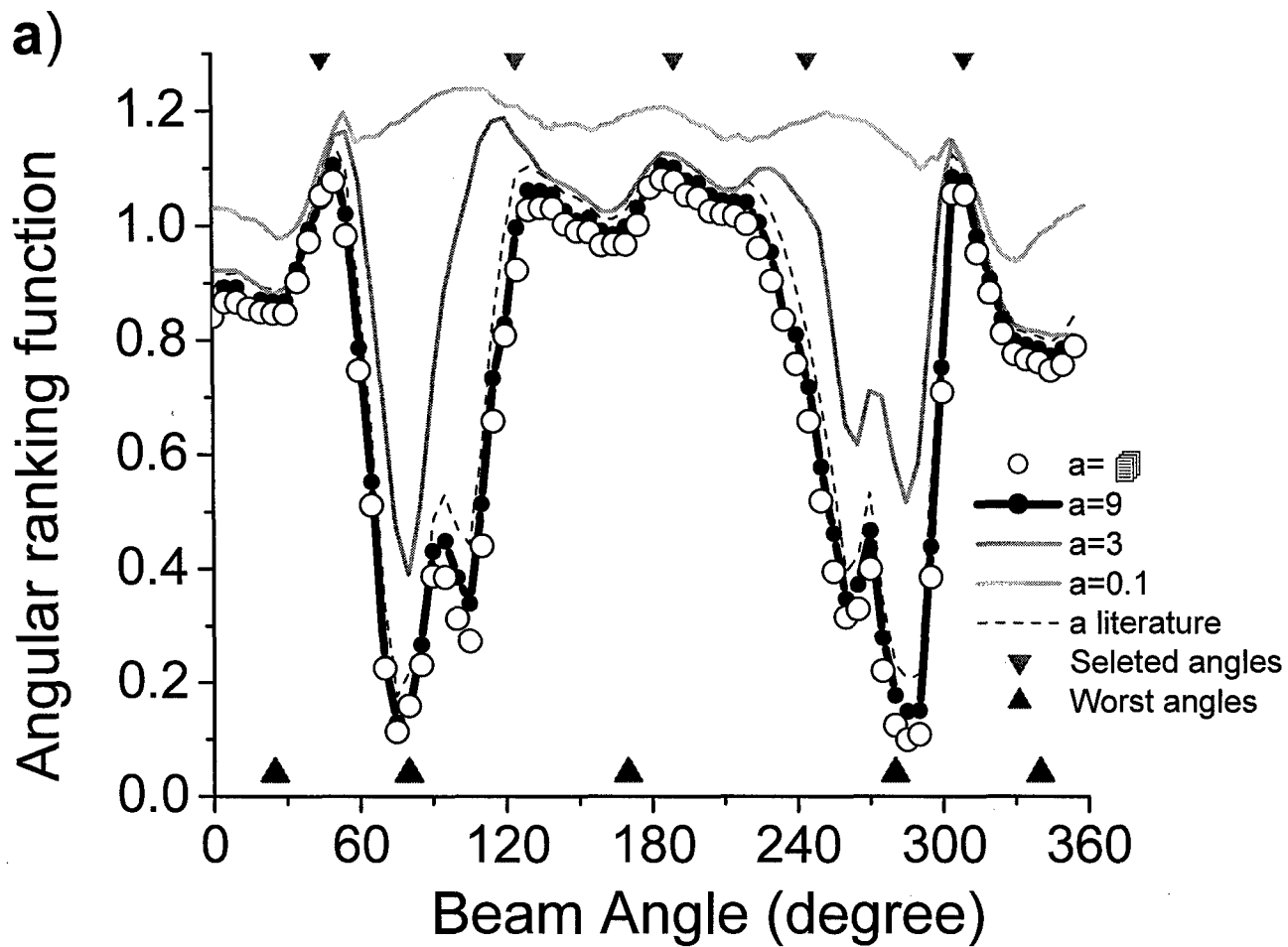
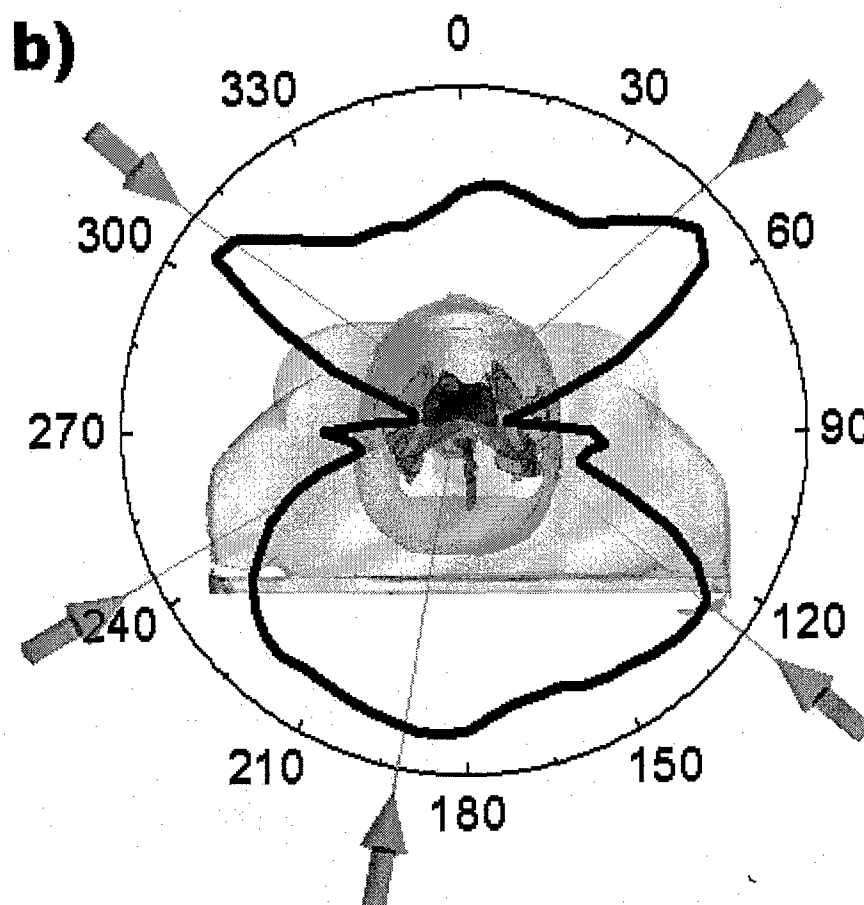


FIGURE 4





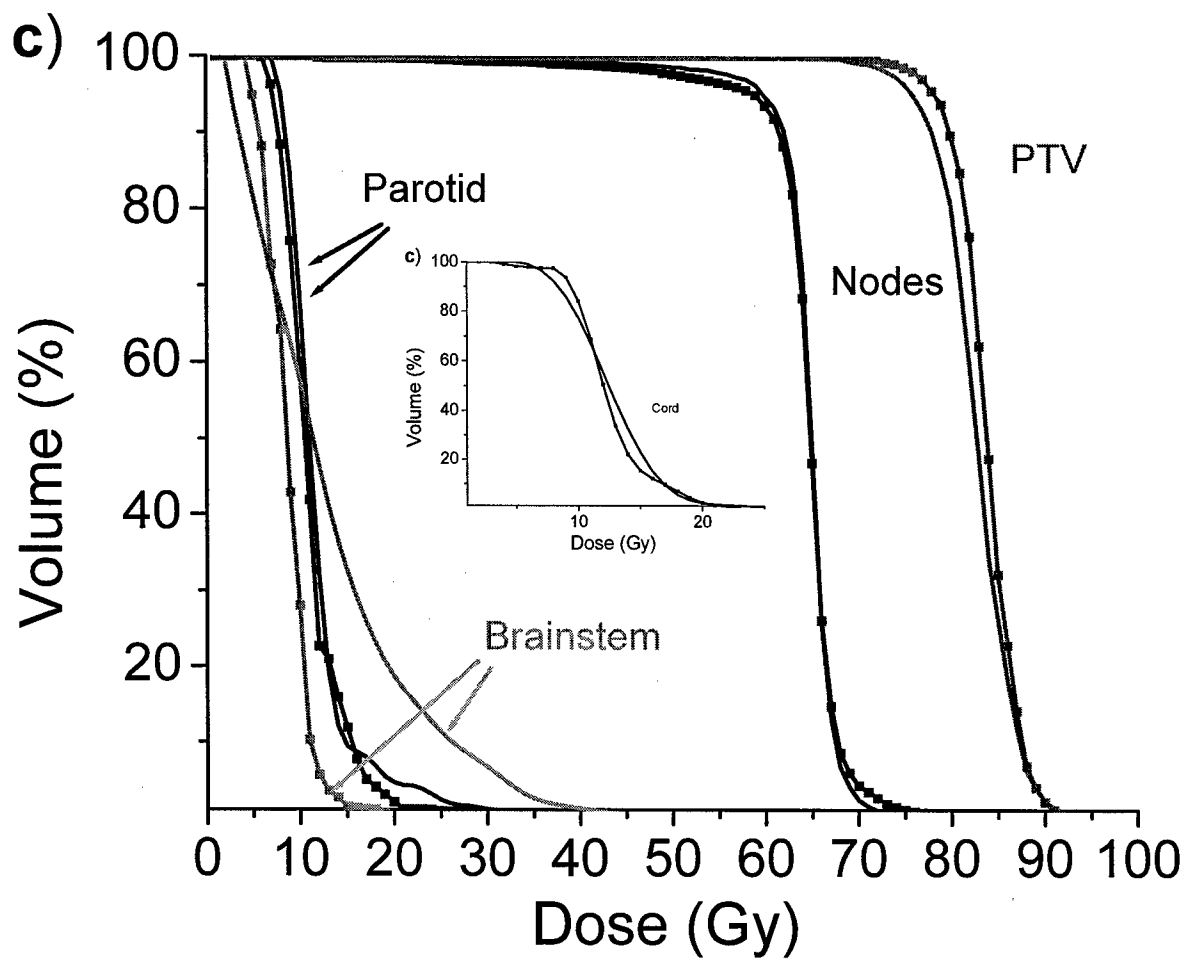


FIGURE 5

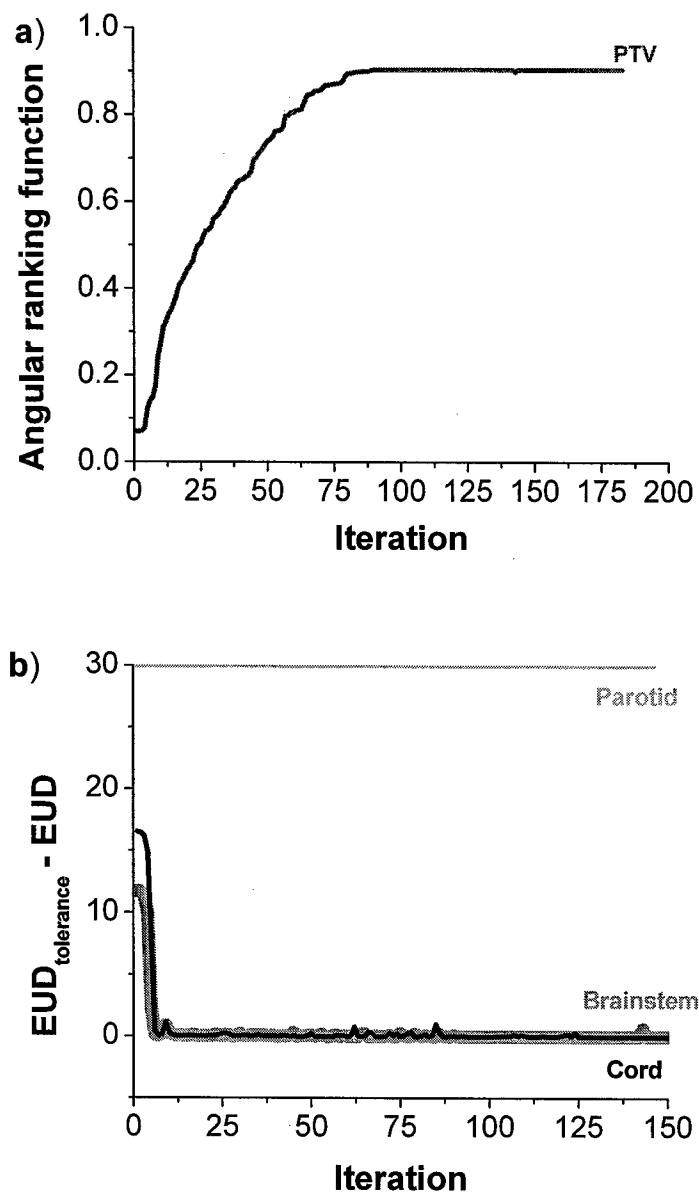


FIGURE 6

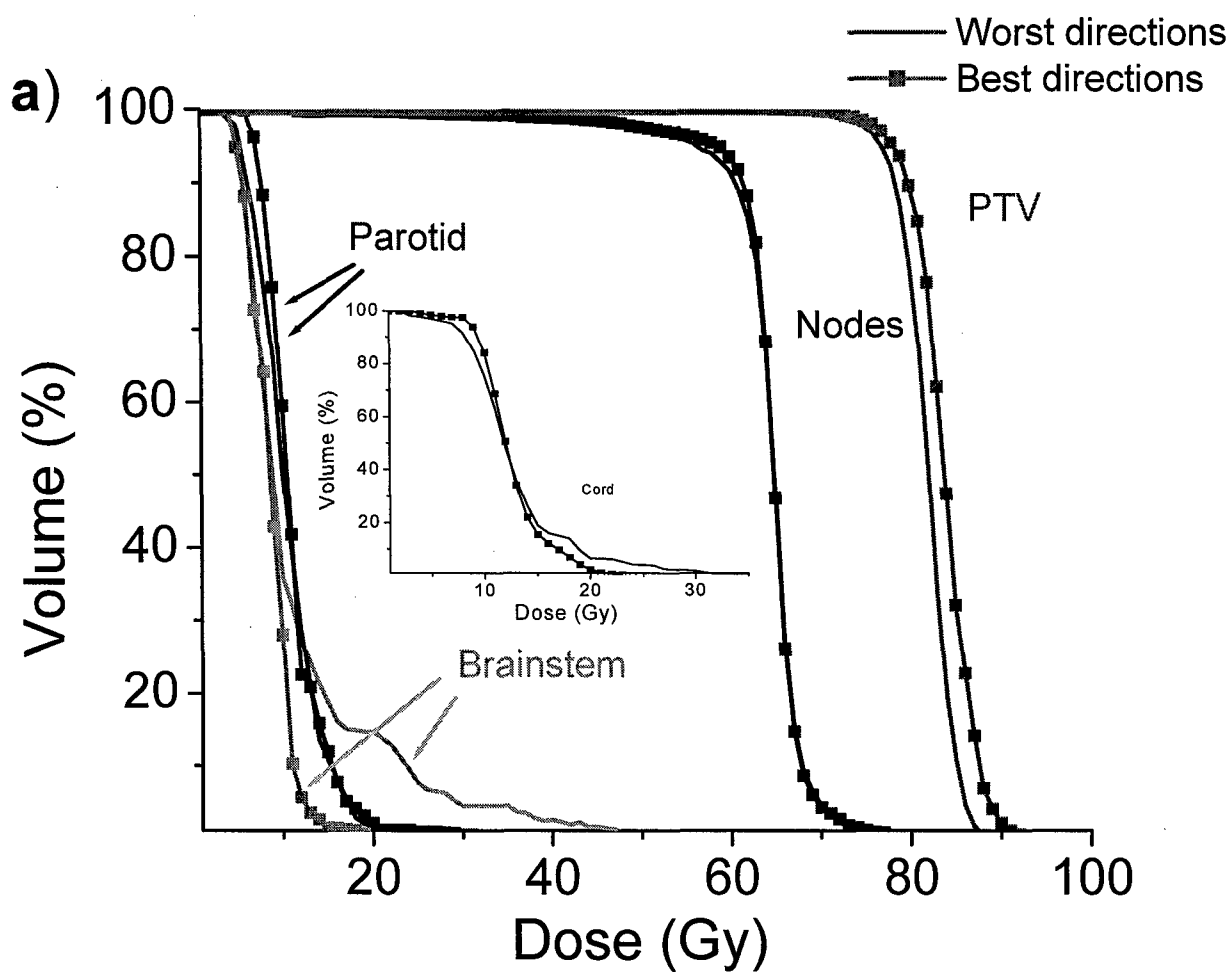


Table 1. EUD tolerances (45) and calculated EUD values for IMRT plans with five equi-angled and BEVD-selected beams for the abdominal case.

	L Kidney	R Kidney	Liver	Cord
<i>a</i>	5.1	5.1	0.59	7.4
EUD tolerance	28	28	39	43
EUD (equispaced beams)	16.82	11.96	8.77	5.31
EUD (BEVD-selected beams)	14.24	8.34	6.97	6.68
EUD (beams with non-EUD-based BEVD)	15.23	9.58	7.03	6.12

Table 2. EUD tolerances (45) and calculated EUD values for IMRT plans with five equi-angled and BEVD-selected beams for the head and neck case.

	Parotid	Brainstem	Cord	Score
<i>a</i>	5.0	4.6	7.4	
EUD tolerance	30	49	43	
EUD (equispaced beams)	16.67	27.55	20.99	1.30
EUD (BEVD-selected beams)	14.00	9.50	19.56	1.33
EUD (beams with worst BEVD scores)	18.71	22.82	20.29	1.27

APPENDIX 9

Image Registration with Auto-Mapped Control Volumes

Eduard Schreibmann, Ph.D. and Lei Xing^{a)}, Ph.D.

Department of Radiation Oncology, Stanford University School of Medicine,
Stanford, California 94305-5847, USA

a) Author to whom correspondence should be addressed.

Stanford University School of Medicine
Department of Radiation Oncology
875 Blake Wilbur Drive
Stanford, CA 94305-5847
E-mail: lei@reyes.stanford.edu
Phone: (650) 498 7896
Fax: (650) 498 4015

Submitted to: International Journal of Radiation Oncology, Biology, Physics

ABSTRACT

Purpose: Many image registration algorithms rely on the use of homologous control points on the two input image sets to be registered. In reality, the interactive identification of the control points on both images is tedious, difficult and often a source of error. The purpose of this work is to automate the selection of control points for both rigid and deformable image registrations and to demonstrate the utility of the new approach by using a few examples.

Materials and Methods: The registration of two images in our approach proceeds in two steps. First, a number of small control regions having distinct anatomical features are identified on the model image. Instead of attempting to find the correspondences of the control regions in the reference image through user interaction, in the proposed method, each of the control regions is mapped to the corresponding parts of the reference image by using an automated image registration algorithm. The conventional automated image registration algorithm is then used to complete the image registration process with the auto-determined control points. A normalized correlation function was used as the metric in both the selection of the control volumes and the final image registration. The deformable registration was modeled by free form deformations based on B-spline. The limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS) was used to optimize the metric function. The performance of the registration-in-registration approach was examined by registering clinical cases of different modalities.

Results: An image registration algorithm with auto-mapped control regions has been implemented for intra- or inter-modality image registration. The convergence of the algorithm was confirmed by starting the registration experiment from over 50 different initial conditions. It is shown that the technique can

match a CT image with an accuracy of ~ 1 mm. Application of the technique to a clinical PET/CT and MRI/CT registrations indicated that it is capable of producing clinically sensible mapping. Similar level of success was found when the method was applied to the deformable registration of two phases in a 4D CT image set.

Conclusions: The proposed method of determining the control points greatly reduces the complexity involved with the determination of homologous control points and allows us to minimize the subjectivity and uncertainty often occurring in the use of user-defined control points. Patient studies have indicated that the two-step registration technique is reliable and provides a valuable tool to facilitate both rigid and non-rigid image registration problem.

Key words: image registration, image fusion, deformable model, BSpline

INTRODUCTION

Imaging is at the foundation of many clinical interventional procedures and development of an effective image registration technique has been one of the most important research areas in imaging science. Generally speaking, registration is to align two imaging data sets in a common coordinate space by transforming the higher resolution data set (the “model”) while keeping the other one (the “reference”) fixed. Depending on the mathematical nature of the transformation, image registration is divided into rigid and deformable registrations. In rigid transformations, it is assumed that the geometry of the object is identical in the two input images and no distortion occurs in the image acquisition process. When working in the “world coordinate” in which the lengths in both images are measured in the same scale, a rigid transformation consists of six degrees of freedom: three displacement parameters and three rotational parameters. Deformable registration, on the other hand, is more complicated and entails the modeling of voxel dependent distortion in addition to the translation and rotation. Clinically, the need for a robust image registration algorithm to compare/fuse images representing the same structures imaged under different conditions or on different modalities is ever increasing because of the extensive use of multi-modality imaging and the emergence of new imaging techniques and methods.

Computer-based image registration has gained widespread popularity in the last decade and is used in routine clinical practice. In this approach, the matching of the two input images, a reference image and a model image, is formulated into an optimization problem and the best registration of the two images is obtained by iteratively comparing various possible matches until no better registration can be found. The search for the optimal match of the input two images is usually gauged by a ranking function constructed based on some physical considerations. Depending on the nature of the input images, the formulation of the problem can be highly complicated and the truly optimal solution may not be readily attainable. To facilitate the computer decision-making process, image pre-processing or

user interaction may be required, especially when dealing with deformable image registration problems. Along this line, a commonly used strategy is to locate several homologous anatomic landmark pairs on the two input images, as it is general true that inclusion of prior system knowledge often leads to a better matching. In order to accurately identify the control points on both images, a prerequisite is that the user must have a detailed understanding of the patient anatomy and the characteristics of the two modalities. The point pairs are usually obtained interactively with the user repetitively exploring the input image sets and each time trying to locate a point in both of them. Due to the 3D nature, the process is rather tedious and difficult to perform. In addition, there is some inaccuracy or subjectivity in the resulting match due to the user dependence in selecting the control points.

The purpose of this work is to develop a general method to facilitate the selection of control points for both rigid and deformable image registrations. Under an assumption that sufficient common features exist in the two input studies, the registration of two images in our approach proceeds in two steps. First, a number of small control regions having distinct anatomical features are identified on the model image in a more or less arbitrary fashion. Instead of attempting to find the correspondences of the regions in the reference image through the judgment of the user, in the proposed method, each of the control regions is mapped to the corresponding part of the reference image by using an automated image registration algorithm. The mapping of the control regions is generally efficient and robust provided that sufficient information is included in each of them. After the mapping, a conventional automated image registration algorithm based on the use of pre-determined control points can be employed to complete the rest of the image registration process. This new way of determining the control points eliminates the need for manual determination of the homologous control points in rigid or deformable registration and allows us to minimize the subjectivity and uncertainty associated with the use of user-defined control points.

METHODS AND MATERIALS

Software platform

All calculations are implemented using an open-source software toolkit named the Insight Toolkit (ITK) (1), which consists of template-based codes for a large number of image visualization, segmentation and registration classes. The programs contained in ITK are easily extendable, making it an ideal platform for the development of image registration methods. Concise and clear descriptions on the use of the available subroutines are provided in the ITK manual.

Selection of control regions on the model image

Image registration is a highly non-trivial optimization problem and sub-optimal or even non-optimal solution may be resulted when it is done without inclusion of any *a priori* knowledge of the system. A method for enhancing the success rate of the calculation is to provide partial guidance to the optimization program by introducing a number of homologous control points identified on the model and reference images at different spatial locations. In general, the point pair should represent the same anatomic location in the two data sets. While the method is inherently efficient because of the use of a priori information of the system, a challenge here is that the spatial arrangements of the reference and model points must be brought into optimal agreement. This is often performed by a skilled user and there exists considerable inter-person variation in the identification of the control point pairs.

Different from the previous approaches, we replace each control point on the model image by a small volume, coined as control volume. A control volume is placed on a location where deformation is negligible (e.g., in or nearby a bony structure). It is preferable that sufficient imaging feature/information is contained within the volume. The underlying assumption here is that, the regions on a model image are generally not equivalent and some regions can be more reliably mapped onto the

reference image than others. The use of “rigid” control region that possesses rich internal intensity pattern makes the mapping of the region from the model image to the reference image simple and unique. The determination of the size of the control region is generally a balance of a few factors. If the volume is too small, there would be no enough structurally unique feature in the region to warrant an accurate mapping of the region. On the other hand, it is difficult to specify a large control volume that meets the general selection criteria of the control volume unless in a special situation where the volume encloses a large portion of bony structure. Typically, the volume is spherical or cubic shape with a dimension of ~1cm, but depending on the application, other shaped volume and dimension should also be acceptable.

Mapping of control regions from the model image to reference image

In figure 1 we show the flow chart of the new image registration scheme. The input to the registration software is the images to be registered: a model image and a reference image, described by their intensity distributions $I_\alpha(\mathbf{x})$ and $I_\beta(\mathbf{x})$, respectively. A virtue of the new approach is that the control regions need to be identified only on the model image in a fairly arbitrary fashion. The correspondences of these regions on the reference image are determined automatically through the use of an automatic image registration algorithm. In a sense, the determination of control regions here is a process of registration within a registration. For intra-modality registration, we use the normalized correlation of control regions between the model points and the corresponding reference points

$$f = \frac{\sum_{i=1} I_\alpha(\mathbf{x}_i) I_\beta(T\mathbf{x}_i)}{\sqrt{\sum_{i=1} I_\alpha^2(\mathbf{x}_i) \sum_{j=1} I_\beta^2(T\mathbf{x}_j)}} \quad (1)$$

as the criterion for assessing the goodness of the mapping of the control regions. In Eq. (1), i and j are the voxel indices within a control region α in the model image, $I_\alpha(\mathbf{x}_i)$ is the intensity of the voxel at

\mathbf{x}_i on the model image α , and $I_\beta(T\mathbf{x}_i)$ is the intensity of the reference image β at where the voxel \mathbf{x}_i is mapped. We note that other type of metrics should be equally applicable to deal with the problem here. For inter-modality registration such as CT and MRI, the commonly used mutual information metric, as was first proposed by Maes and Viola (2-4), can be used as the matching criterion for the mapping of the control regions.

The above mapping directly uses the value of all pixels within a control volume as the information feature with the assumptions that both data sets show the control volume similarly and that the control volume pair is related by a rigid transformation. An idealized situation is that the intensity of each voxel in a control region in the model image differs from its corresponding reference pixel by a factor. In practice, the two data sets may be affected by noise, sampling, and clinical or experimental situation. Therefore the mapping of a control volume is considered optimal if the above normalized correlation function is minimal or mutual information is maximal. Numerous optimization algorithms can be employed to minimize the matching criteria. In our calculation, the control volumes are mapped sequentially and independently. For a chosen control region, an iterative optimization algorithm is used to update the six transformation parameters until no further improvement can be achieved. The iterative calculation is outlined in the next section. After one volume pair is finished, the calculation moves to the next control region and the iterative calculation is repeated.

Optimization of the NCC function

The NCC given in Eq. (1) or the mutual information metric for a given control volume is a function of its translational and rotational parameters. For each control volume, the optimization of the function with respect to these parameters is done by using the limited memory *BFGS* algorithm, referred to as *L-BFGS* (5). The detail of this algorithm has been presented in Refs. (6, 7) and will not be repeated here. During the mapping of a control volume, the iterative calculation based on *L-BFGS*

algorithm continues until a pre-set maximum number of iterations (typically, 100 iterations) is reached. The convergence behavior of the *L-BFGS* algorithm is studied by starting the registration calculation with different random initial positions.

For rigid registration, a single control volume is, in principle, sufficient to derive the translational and rotational parameters. To be robust, we typically select two to three control volumes and the final registration parameters are determined by averaging that derived from the mapping of each individual volume. The use of control volume for rigid image registration is especially useful when there are artifacts in one or both input images. For deformable registration, we usually select four to six control regions and the translation and rotation of each volume are generally different. The detail of the calculation is described as follows.

BSpline deformable registration with incorporation of the mapped control volume information

In BSpline deformable image registration calculation, the image is divided into a grid with N^3 cells. The corner of a lattice cell is referred to as a node and is indexed by i ($i=1, 2, \dots, N^3$). The displacement of a node i is specified by a vector \mathbf{x}_i and the displacement vectors, $\{\mathbf{x}_i\}$, of a collection of nodes characterize the tissue deformation. The displacement at a location \mathbf{x} on the image is deduced by fitting a polynomial expressed using the basis spline (BSpline) (8, 9) to the grid nodes \mathbf{x}_i . Unlike other spline models, the BSplines are locally controlled. That is, the displacement of an interpolation point is influenced only by that of the closest grid points and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations.

Mathematically, the task of a non-rigid image registration is to find the transformation matrix, $T(\mathbf{x})$, that maps an arbitrary point \mathbf{x} on the model image α to the corresponding point \mathbf{x}' on the reference image β (or *vice versa*) in such a way that the best possible match, as measured by the registration metric, is achieved. As stated earlier, a registration calculation can be divided into two

steps: mapping of the control volumes and registration of the two images. The first step is rigid registration in nature and yields the translational and rotational parameters for each control volume. To complete the registration of the two images, we implement a BSpline deformable registration model, in which the matrix coefficients of $T(\mathbf{x})$ are the node displacements and are adjusted iteratively to minimize the NCC function or maximize a mutual information metric. The information derived from the first step is incorporated into the general deformable registration by intentionally assigning the control volumes as the control volumes are usually identified as some of the nodes. With the two-step registration, some of the matrix elements in $T(\mathbf{x})$ are known with high confidence after the mapping of the control volumes. This greatly facilitates the determination of the rest matrix elements or the rest nodal displacements. We compute these displacements by optimizing the NCC function (1) with respect to the displacements of the remaining nodes, $\{\mathbf{x}_i\}$. Once again, the *L-BFGS* (5) is used to optimize the system here.

Search space characteristics and convergence analysis

In image registration, the change of the metric function as a function of the relative displacement of the two input images provides a useful evaluation of the search space. It is intuitively conceivable that the value of the metric function is minimum/maximum when two images are aligned and increases as they moved away from each other. A desirable metric function should exhibit a single distinct global minimum and vary smoothly as the two images become gradually mismatched. For the rigid registration, we plot the metric as a function of relative displacement of the control volume. The transformation parameters in this situation are determined entirely by the mapping of the control volume(s). For comparison, we also plot the function for a conventional approach without the use of control volume(s). In this case, the two input images are moved away each other since the final transformation parameters are obtained based on iterative adjustment of the relative positioning of the

two images. For non-rigid registration, the behavior of each control volume is different and the metric function for each control volume is plotted when as the volume is displaced from its ideally mapped location.

Global convergence, which states that the optimization result should be independent of the initial values of the system variables, is a desirable property of an optimization algorithm. The convergence behavior of the proposed two-step method is studied for both rigid and non-rigid image registrations by starting the registration calculation with 50 different random sets of initial transformation parameters. In each case studied in this work and for a given initial mismatch of the input images, we plot the metric value as a function of iteration step. The results are compared with that obtained using conventional approach when treating each input image as an entity. In addition, the shifts in x-, y-, and z-direction resulted from each of the 50 calculations is plotted for the two-step registration and the conventional registration.

A checkerboard display tool, where images to be compared are merged together in a chess-like pattern, is used to illustrate the difference between the two input images before and after registration. At the transition zone it is possible to measure displacements between anatomical structures. This method validates the goodness of a registration. Ideally, when the two images are perfectly aligned, no differences should be visible in the checkerboard of the two input image sets.

Case studies

The utility of the proposed registration-in-registration method is demonstrated by studying two rigid registrations and a non-rigid registration case. The first one is a rectal cancer patient who has undergone both CT and [^{18}F] fluorodeoxythymidine (FLT)-PET scans. Here FLT is a marker for thymidine kinase activity, representing one of the potentially more selective tracers that have been under intense investigations. A side feature of FLT-PET is that bony structures appear clearly because

of the high uptake of hemotopoitic cells within the bone matrix, which makes the registration of PET and other anatomic imaging modalities simple. The registration-in-registration technique described above is used to fuse the CT and FLT-PET images and the results are compared with that obtained using conventional approach without relying on the use of control volumes. For this case we use a three rectangular volumes of 7 cm.

The second case consists of the registration of CT and MRI images for a brain tumor case. Although the head is commonly believed to have no deformation, in practice small differences may exist between images acquired using different modalities or under different conditions. The patient studied here uses an immobilization mask in the CT scan but not in the MRI. Additionally, this patient has a tumor in his eye, leading to different MRI values as compared to the normal MRI in this region.

The input in the deformable image registration study is the CT images acquired at the expiration and inspiration phases using a 4D CT protocol. A set of control volumes is selected on the exhale image (figure 7c) and a successive rigid registration is carried out to find their locations in the inhale image. These homologous control volumes pairs are used to evaluate the nodes values of the BSpline deformable algorithm. The registration calculation then proceeds to the next step to match voxels on the exhale image with that of inhale image by using the BSpline method as described earlier.

RESULTS

Case 1: Rigid registration of CT and FLT-PET images for an abdomen case

The input CT and FLT-PET images used for this study are shown in the first and second rows of figure 2. The results obtained with the control volume-based registration are shown in the third and

forth rows. The control volumes relative to patient geometry is depicted in the right of the forth row of figure 2 as a red wire frame. To visually evaluate the registration results, in the third row of figure 2 we show a checkerboard comparison of the PET and CT images after registration. The sagittal checkerboard image of the PET and CT is shown in the left of the forth row of figure 2. The images seem to be registered very accurately, with no geometric mismatch visible in the transition zone. A 3D view of the registration is presented in the right of the forth row of figure 2, where an excellent coincidence is observed between the bony structures revealed in the CT image (white) and the PET images (orange). In both 3D rendering and checkerboard display, the bladder is bright in the PET image, but barely visible in the CT images. The feature inherent to a certain type of images but not others is problematic in multi-modality image registration since it “disturbs” the behavior of the metric function. Another commonly seen example of this is the registration of CT images containing metal artifacts with MRI data (see next section), in which no artifacts present. Control volume(s)-based registration permits us to exclude the unphysical regions based on *a priori* knowledge and provides an effective way to find the truly optimal solution in this situation.

In figure 3 we compare the NCC metric space when the whole image is used with that when a control volumes placed in the lumbar region and femur heads are used. The surface plot in figure 3 represents the NCC function values when the two input images are shifted away from each other in different directions and with different displacements. The function values are color-coded from red, representing sub-optimal matching, to blue, representing the optimal registration sought after. In the former case, it is seen that the metric function is not smooth and there exist multiple local minima. Aside from the fact that the PET image is generally noisy, the high-intensity voxels in the bladder (caused by the rapid excretion of FLT through the urinary system) that do not have a correspondence in the CT image also influence the convergence in the former case. The minimum in the control volume-based metric is, on the other hand, clearly visible and the search space is very smooth. We note that a

single minimum is present. Moreover, the NCC value at the minimum in this case is 0.22, which is two orders of magnitude deeper than that when the whole image is used for registration. The improvement in the functional behavior of the metric makes the image registration calculation much more efficient. This statement is also supported by the following analysis.

To examine the algorithm's ability to find the correct shifts to the best match, we start the registration from different initial values of the transform. In figures 4a we show the NCC as a function of iteration step when starting from 50 randomly assigned initial trial transformation parameters. While the trajectory for each optimization calculation is different, they all converge to the same NCC value in the control volume-based calculations. The convergence of the calculation is also evident from figure 4c, where it is shown that, for the given mismatch of the two input images, all 50 calculations with different initial trial transformation parameters lead to the same shift values in the three directions along x-, y-, and z-axis of 1.75mm, 3.5mm, and 5.4mm respectively. This is, however, not the case for the conventional approach based on the information contained in the whole image entity. As seen from figure 4b, the NCC converges to different values when different initial conditions are used. Furthermore, the shifts (figure 4d) resulted from the calculation fluctuate from calculation to calculation and represent, at best, sub-optimal solutions most of the times. We notice that the miss-registration in the z-direction is particularly serious. The results are not surprising because the PET data are quite noisy and contain structures having no correspondences in CT images.

Case 2: Rigid registration of CT and MRI images for a brain case

In figure 6 we compare the control volume-based and conventional MI based registration of CT and MRI data at four different slice locations for a brain case. Although the head is believed to be an object with no deformations, in practice, difference in CT and MRI may appear in the two input data sets to adversely affect the registration. For example, the patient analyzed uses an immobilization mask

during CT scan but not in the MRI imaging. There is a tumor in the right eye of the patient, which makes the MRI appear differently from a regular situation. Additionally, CT artifacts exist in the mandible region due to the denture, which does not have much impact on the MRI data. These factors complicate the registration algorithm in analogous to the bladder artifact does to the previous patient.

Three control cubic volumes are placed in the model image. In general, a MI metric is preferable for multi-modality image registration for CT-MRI registration. The control volume-based registration method outperforms the conventional approach in both computational speed and the quality of the registration. Indeed, the computing time for the two different techniques to find their "optimal" solutions is reduced from hours for the conventional approach to minutes for the control-volume-based calculation. In the skull region, there is a mismatch between CT and MRI images in the conventional MI registration with the whole images because of the presence of immobilization mask in the CT image. This mismatch is eliminated in the control volume based calculation. The improvement is more pronounced in the slices passing through the eyes of the patient. The presence of the eye tumor in the MRI confuses the whole-image based MI registration. As a result, the algorithm tries to correct the non-exist problem and yields an unrealistic registration of the CT and MRI in this region. As seen from figure 5, in the standard registration the tumor defined in MRI is dragged down from its ideal location (slice 2) to accommodate the artifacts in slices 1 and 4. The control volume-base calculation also avoided the disturbance of the CT artifacts in the mandible resulted from the denture.

For a given mismatch of the two input images, we have run 50 calculations with different initial trial transformation parameters. Similar to the case shown before, the control-volume-based algorithm yields the same shift values in the three directions along x-, y-, and z-axis. For the conventional approach based on the information contained in the whole image entity, we find that the final transformation parameters vary from test to test and does not converge to the same solution, even though they all fall into a certain range of fluctuations. Consequently, the shifts (figure 4d) derived from

conventional calculation represent, at best, sub-optimal solutions. We notice that the miss-registration in the z-direction is particularly serious. On the contrary, the control- volume-based calculation yields consistent patient shift parameters in all 50 calculations. This confirms the conclusion drawn in the previous case study.

Case 3: Deformable registration of exhale and inhale CT images for a thorax patient

The proposed method is applicable to both rigid and non-rigid image registration problems. The inputs to be registered in this example are CT images acquired at expiration and inspiration (figures 7 a and b). A few control volumes are selected on the model image (figure 7c) and they are then mapped to find their corresponding locations in the reference image (figure 7d).

Similar to the convergence analysis presented in the first example, in figure 8 we plot the NCC metric for three of the five control volumes when these volumes are shifted away from their best match position in different directions and with different displacements. To better understand the system, we have chosen two different sized control volumes: 20 mm (upper row) and 40 mm (lower row). In both situations, the metric space is smooth and has a very pronounced ridge. A zoom of the ridge region presented as inset in figure 8 shows that the true minimum is very close to the other minima in the ridge. For example, for the control volume of 20 mm, the global minimum has a value of -0.91, while the neighbor ridge has a value of -0.90. The legend in each inset shows the small range metric variation along the ridge. This behavior is resulted from the fact that the intensity variation within the selected control volume is predominantly in the direction perpendicular to the boundary of the lung. When comparing the metric function for different sized control volumes, we note that the search space characteristics do not change significantly. For the size of 20 mm, the peaks and valleys of the search space are more pronounced because, as the number of pixels reduces, a small miss-match would lead to a large variation in the cost function. As the control volume size increases, the details of the cost

function are “evened out ” by the large number of voxels. In both situations, the positions of the minimum are found to be $(x, y, z)=(-6\text{mm}, -12\text{mm}, 4\text{mm})$, $(-2\text{mm}, -12\text{mm}, 3\text{mm})$ and $(-13\text{mm}, 2\text{mm}, 4\text{mm})$ for the three control volumes, respectively. In practice, determining the size of a control volume is a matter of tradeoff between a few factors, such as the accuracy and sensitivity against a small mismatch of the control volumes, robustness, and calculation speed.

After the mapping of the selected control volumes, a warping using the spline model is used to register the whole image with the control volumes serving as the nodes of the BSpline calculation. The results of the deformable registration are shown in figure 7. The model image is represented as semi-transparent background in figure 7c. The selected rectangular control volumes are also represented in figure 7c overlays on the background image, with the image in each control volume cropped from the reference image to assess differences between the two input images to be registered. After registration, no difference between the mapped model image and the reference image is observable (figure 7d). The results obtained with the control volume-based method are compared with that obtained using a “brute-force” BSpline formalism (6). Moreover, it removes the subjectivity and potential errors that are often associated with the manual landmark based approach. In figure 9a we show the checkerboard image of the two input images as given in figures 7a and 7b. Before registration, deformations up to 1 cm in the two phases are clearly visible in the checkerboard transition zone and are marked with red arrows. Both control volume-based and the BSpline models (figure 9b) were able to register the two images. The mapped inhale images from the two different approaches are shown in figure 9b using a checkerboard tool. The difference between the two mapped images was found to be less than 2 mm in any region. We noticed that the control volume-based registration takes only a few minutes to complete the process, while the time for the BSpline registration a few hours. The significant improvement in computational speed is attributed to the fact that the BSpline model needs to optimize 3375 variables and evaluate the metric over 23 millions pixels, which is orders of magnitude beyond the data

manipulations that the control volume based approach has to handle. To compare, we remind that only six transformation parameters are required for each control volume and, because the positions of these control volumes or nodes are known after the first step of mapping, the remaining calculation for deformable registration can be accomplished with fewer iterations.

The convergence behavior of the deformable registration is assessed in a similar fashion as described previously. We repeated the registration calculation 100 times with randomly assigned initial transformation parameters for the control volumes. The initial displacement of a control volume was generated randomly (in a range of 1 cm). The transformation parameters of the first three control volumes in x-, y-, and z-direction for the 100 independent registration calculations are plotted in figure 10. For the first control volume, shifts in x-, y-, and z-direction were found to be (-6, -12, -4.). The transformation parameters in the three directions resulted from the individual mappings were within a range of 0.90, 1.27 and 1.53 mm, respectively. Similar convergence was achieved for the second and the third control volumes. The variation ranges for the two regions were found to be (1.39, 1.93 1.32) and (1.09 1.55 1.45), respectively.

DISCUSSION

With the recent advancements in image-guided radiation therapy (IGRT), multimodality imaging becomes increasingly important. Full realization of the potential of IGRT and highly conformal IMRT would be impossible without a robust and efficient image registration technique. In general, image registration has two important aspects: formulation of the problem and optimization of the transformation parameters required to match the two input images. In this work we proposed a general two-step registration technique, in which a homologous pair of control volumes are obtained using an auto-mapping algorithm and the pairs are then used as the *a priori* knowledge of the system to facilitate

the registration process. In particular, we present our experience with the selection of model parameters, optimization algorithm, and the validation of the technique. The development of the proposed technique was motivated by an intuitive observation that, there are typically two types of regions in the model image: regions whose correspondence in the reference images is easily identifiable and regions whose mapping from the model image to the reference image is less obvious for a variety of practical reasons, such as tissue deformation, breathing movement, lack of distinct image features, image artifacts, or inherent difference in the imaging modality. Mapping of the two types of regions should be treated differently, rather than equally as implemented in current image registration algorithms. The use of control volumes placed on the reliable regions permits us to eliminate the uncertainty and/or disturbance arising from the less reliable regions, thus improving the performance and robustness of the image registration calculation.

In the case of rigid registration, the final transformation parameters are an average of the parameters derived from the chosen control volumes. The approach is superior over the existing registration technique based on the auto-matching of the two whole image sets because it is less prone to local disturbances from imaging artifacts. In reality, differences in image features due to the use of different acquisition protocols may constitute "noise" to the registration algorithm. Using control volumes provides an effective way to avoid the disturbance and allows us to fully utilize the features that are known with high confidence to more effectively deal with the image registration problem. The two-step registration technique has the advantage of the landmark-based registration without the need of having to go through the hassle of interactive selecting the homologous control point pairs. It is well recognized that interactive selection of control point pairs in landmark-based registration is often subjective and the result may vary from user to user. The proposed registration technique offers operator independence and reproducibility. Of course, it must be ensured that the registrations found by

any automatic algorithm are also reasonable. West et al. (2) noted that careful visual inspection of the results obtained from automatic registrations is crucial in a clinical setting.

For deformable registration, the benefit of the two-step registration is even more significant. In addition to the increased robustness and confidence level, the technique greatly speeds up the calculation process. In conventional BSpline method, for example, the mapping of the nodes or the transformation of the nodes from the model image to the reference image is accomplished by a searching algorithm, which treats the transformation parameters of the nodes as inter-related system variables and hence necessitates an iterative examination of a huge number of trial parameters under the guidance of the registration metric. The amount of computation is proportional to $O(N^3)$, where N is the number of the nodes. In the proposed two-step registration, the selected control volumes can be regarded as BSpline nodes. Therefore, the determination of most, if not all, nodes can be done independently, and this essentially reduces the calculation into a $O(N)$ problem. The reduction in computational time may have important practical implication as the radiation therapy is moving toward more sophisticated image-guided patient localization and dose delivery.

Our study has also shown that the use of control volumes can greatly simplify the optimization calculation incurred in the image registration problem. Over the years, powerful but computationally intensive algorithms, which are capable of statistically escaping from local minima, such as simulated annealing or genetic algorithms, have often been used to attack the optimization problem. Even with the stochastic algorithms, finding the true solution of the problem can still be challenging because of the wide variety of possible pixel intensities within an organ and the complex behavior of the metric function. This has been clearly demonstrated in the examples presented above. The distinct feature of the control volume smoothens the metric function space and makes it simple for the search algorithm to converge to the global solution.

CONCLUSIONS

We have proposed a novel method for auto-selection of the control volumes and described a two-step registration technique for both rigid and deformable image registration. Instead of relying on interactive selection of homologous control point pair on both model and reference images, in the proposed approach the user needs only to identify a few small control volumes on the model image in a somewhat arbitrary fashion, which effectively eliminates the nuisance of the control point pair selection and removes a potential source of error in registration. The method was applied to both rigid and non-rigid image registration problems and our results indicated that the registration is reliable and provides a valuable tool for intra- or inter-modality image registration. The added confidence in the registration and increased speed of calculation, especially in the case of the deformable registration, are also important features of the new technique. Compared to the manual rigid registration, this method eliminates the step of interactively selecting control points, which can be labor intensive and difficult. Compared to the automated method, the method is more intuitive and robust, especially in the presence of image artifacts. Thus the method should have find useful applications in radiation oncology and other image-guided interventions.

ACKNOWLEDGEMENTS

We would like to thank Drs. T. Li, Y. Yang, B. Thorndyke, A. Koong, and B. Loo for useful discussions. This work is supported in part by a research grant from the prostate cancer research program of U.S. Department of Defense (DAMD17-03-1-0023) and the National Cancer Institute (1 R01 CA98523-01).

REFERENCES

1. Ibanez L, Schroeder W, Ng L. ITK Software Guide. Kitware Inc. 2003.
2. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 1997;16:187-98.
3. Wells WM, 3rd, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. *Med Image Anal* 1996;1:35-51.
4. Viola P, Wells III WM. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* 1997;24:137-154.
5. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 1989;45:503-528.
6. Mattes D, Haynor RD, Vesselle H, Lewellen KT, Eubank W. PET-CT image registration in the chest using free-form deformations. *IEEE Transactions Medical Imaging* 2003;22:120-128.
7. Schreiber E, Xing L. Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies. *International Journal of Radiation Oncology Biology Physics* 2005;62:595-605.
8. Lee S, Wolberg G, Chwa KY, Shin SY. Image metamorphosis with scattered feature constraints. *IEEE Transaction Visualization Computer Graphics* 1996;2:337-354.
9. Lee S, Wolberg G, Shin SY. Scattered data interpolation with multilevel B-splines. *IEEE Transactions Visualization Computer Graphics.* 1997;3:228-244.

FIGURE CAPTIONS

Figure 1. Flow chart of the proposed registration procedure. The control volumes are selected by the user only on the model image. A rigid registration algorithm automatically maps the selected control volumes to their corresponding locations on the reference image.

Figure 2. Sagittal, coronal and axial views of the FLT-PET images (first row) and CT images (second row). The checkerboard of the CT and FLT-PET images after registration is shown in the third row and the left panel of the forth row. The right panel shows a stereotactic view of the matched PET and CT images after registration.

Figure 3. Metric space of FLT-PET and CT registration when the whole image sets are used (a) and when three control volumes on the bony structure is used (b). The metric function in control volume-based registration is much smoother

Figure 4 Converge analysis for a rigid registration case with (left column) and without (right column) a control volumes. The control-volume based registration converges to the same cost function value and leads to reproducible shift values in x-, y-, and z-directions (within less than 1 mm). In the standard registration, large variations are observed, as the algorithm is often trapped in local minima.

Figure 5. Comparison of rigid registration case with the newly proposed control volume-based calculation (left column) and conventional whole image-based calculation (right column). In conventional registration, the metric tries to match all the voxel intensities and thus it is more likely to fail in the presence of image artifacts or other noises. In the right row, the images obtained by the conventional procedure are better aligned at the extremities where the artifacts are present, but fail in the regions of clinical interest (such as the slice containing the eye tumor). Using a control volume on the bony anatomy eliminates artifacts and produces better registration.

Figure 6 Converge analysis for a rigid brain CT-MRI registration for the control-volume based registration. All 50 calculations with different initial transformation parameters converge to the same shift values in x-, y-, and z-directions (within less than 1 mm).

Figure 7. The model (a) and reference (b) images for the deformable registration study. Rectangular control volumes, #1 to #5, are placed in the model image and their mappings are shown on the reference image. The anatomy in the regions affected by the respiration does not match initially (c). After registration, the anatomy matches very well, as confirmed by the overlay image (d).

Figure 8. Metric function for the first three control volumes for two different rectangular sizes (20 mm for upper row and 40 mm for the lower row). In both cases, the search spaces are smooth, with a very pronounced ridge. The search space characteristics and the location of the global minima do not depend significantly on the control volume size.

Figure 9. Checkerboard comparison of the model and reference images before (a) and after (b) control volume-based deformable registration. Displacement up to 1 cm is visible before registration, as marked by the arrows in (a). The control volume-based algorithm yielded the virtually the same result as that of the conventional BSpline registration, but with a significant reduction in computational time. A checkerboard comparison of the mapped model images from the two algorithms is shown in (c).

Figure 10. Convergence analysis for the first three control volumes. Presented are shifts of the final transformation as obtained in 100 registration tests starting from different initial transformation parameters. All calculations converged to the same solution to within a range of less than 2 mm.

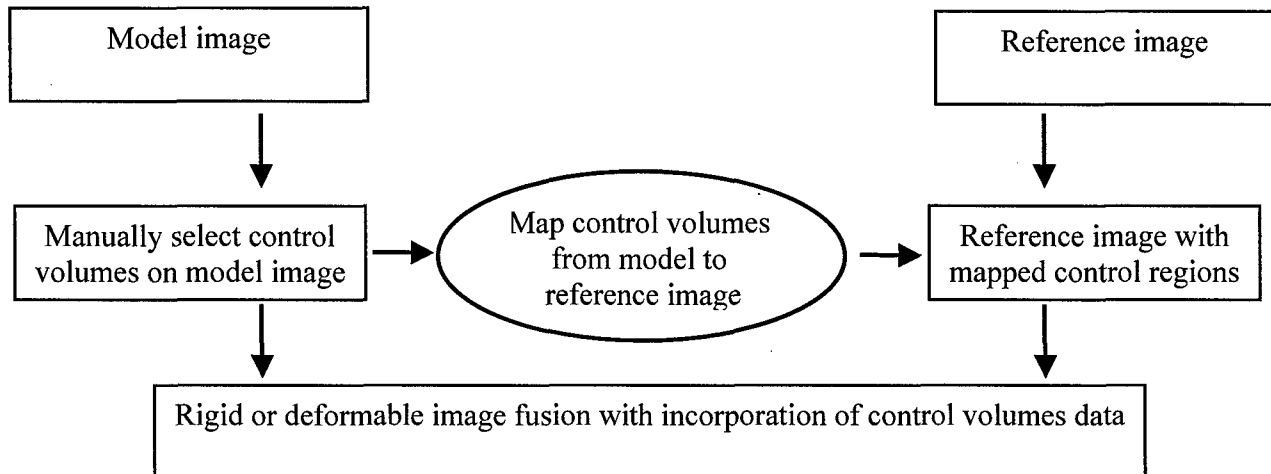




Figure 2

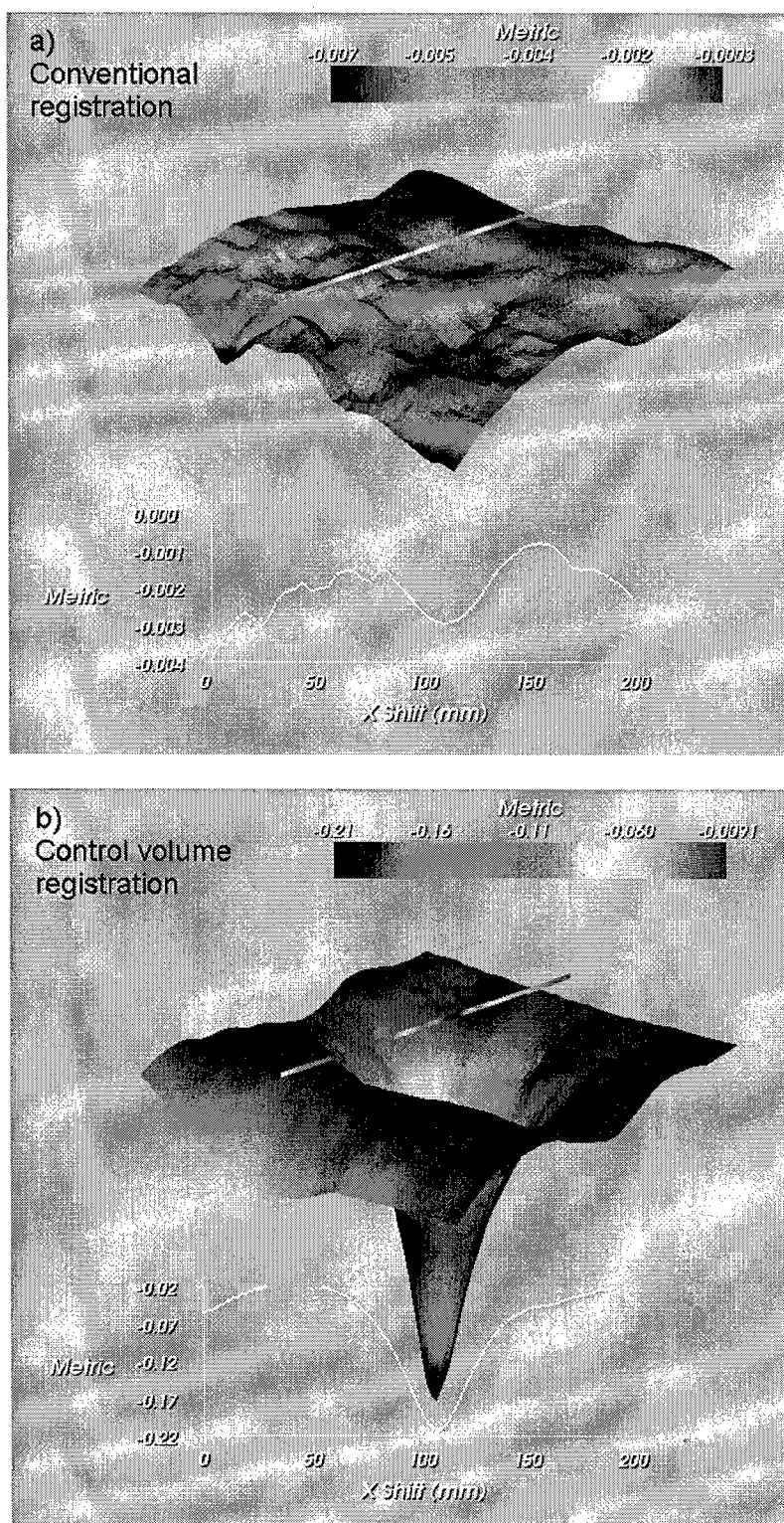
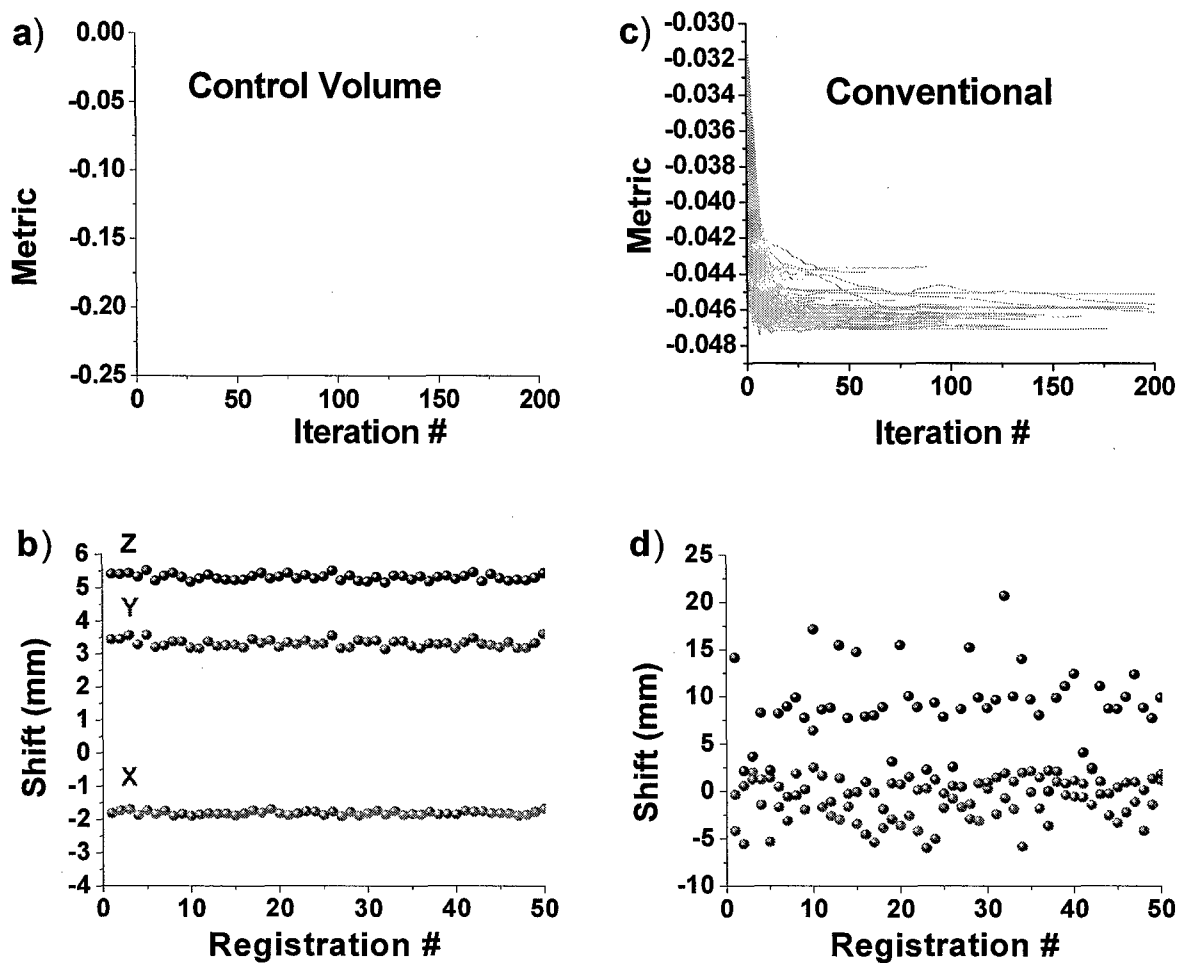


Figure 3



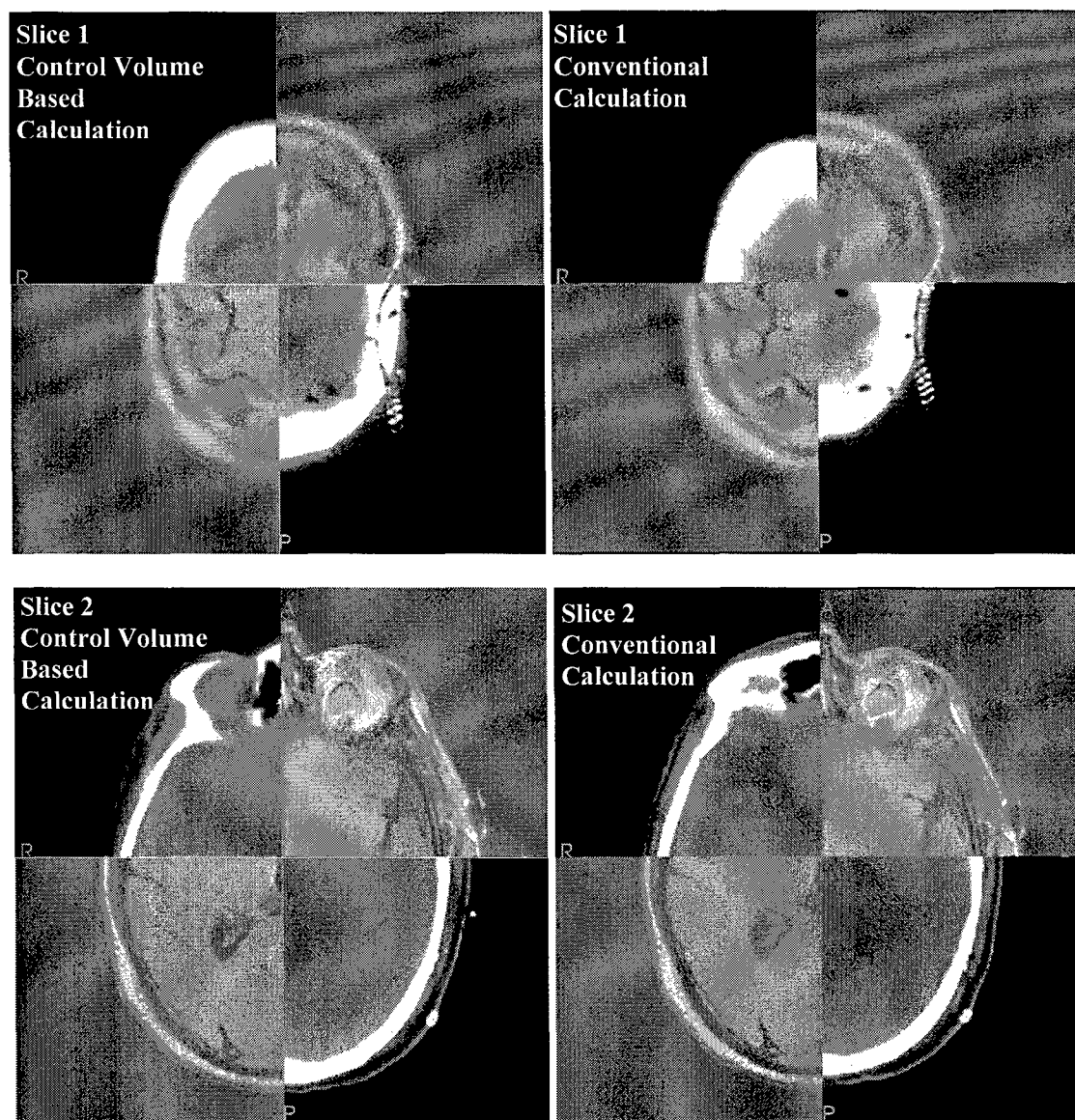
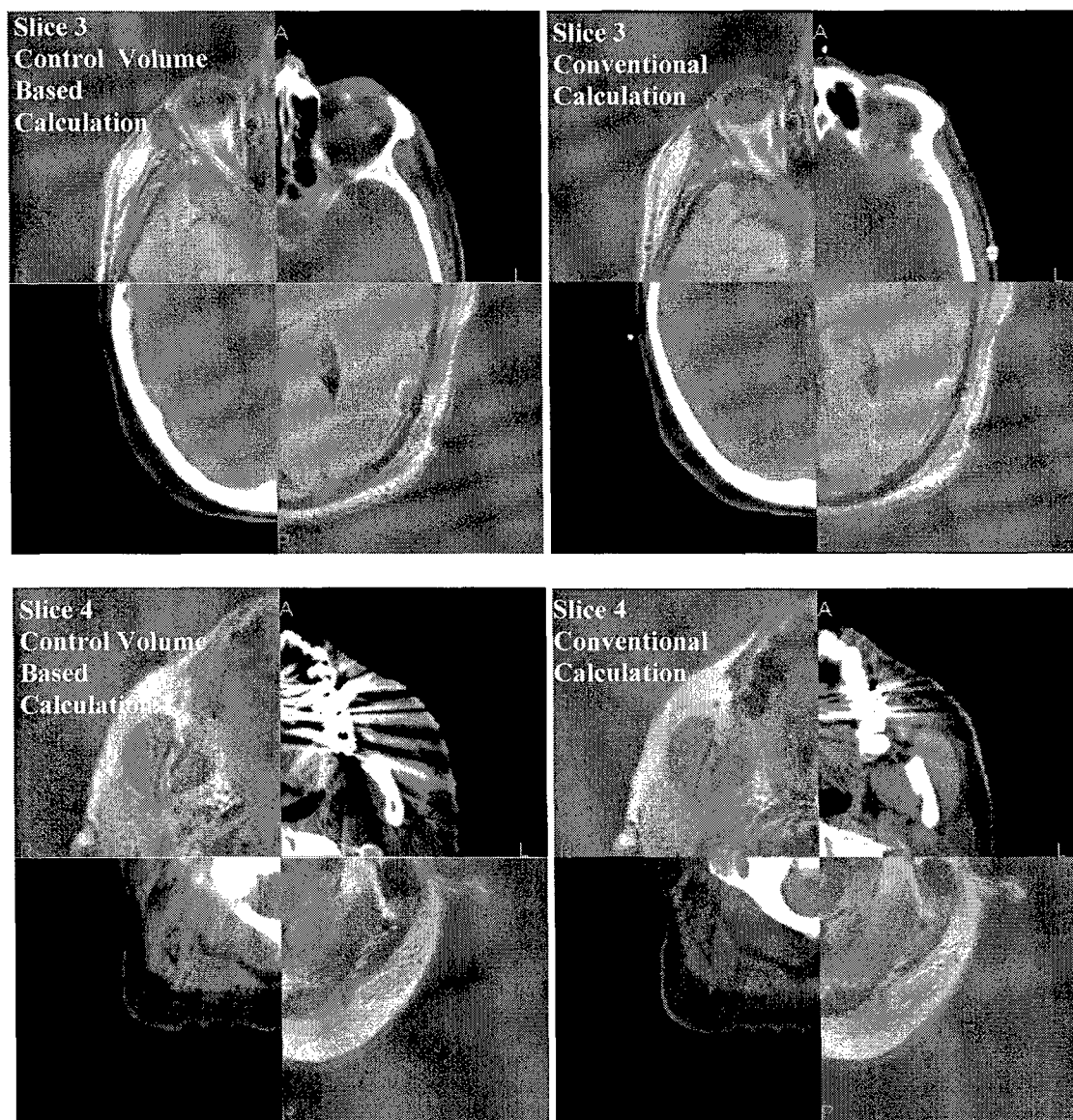
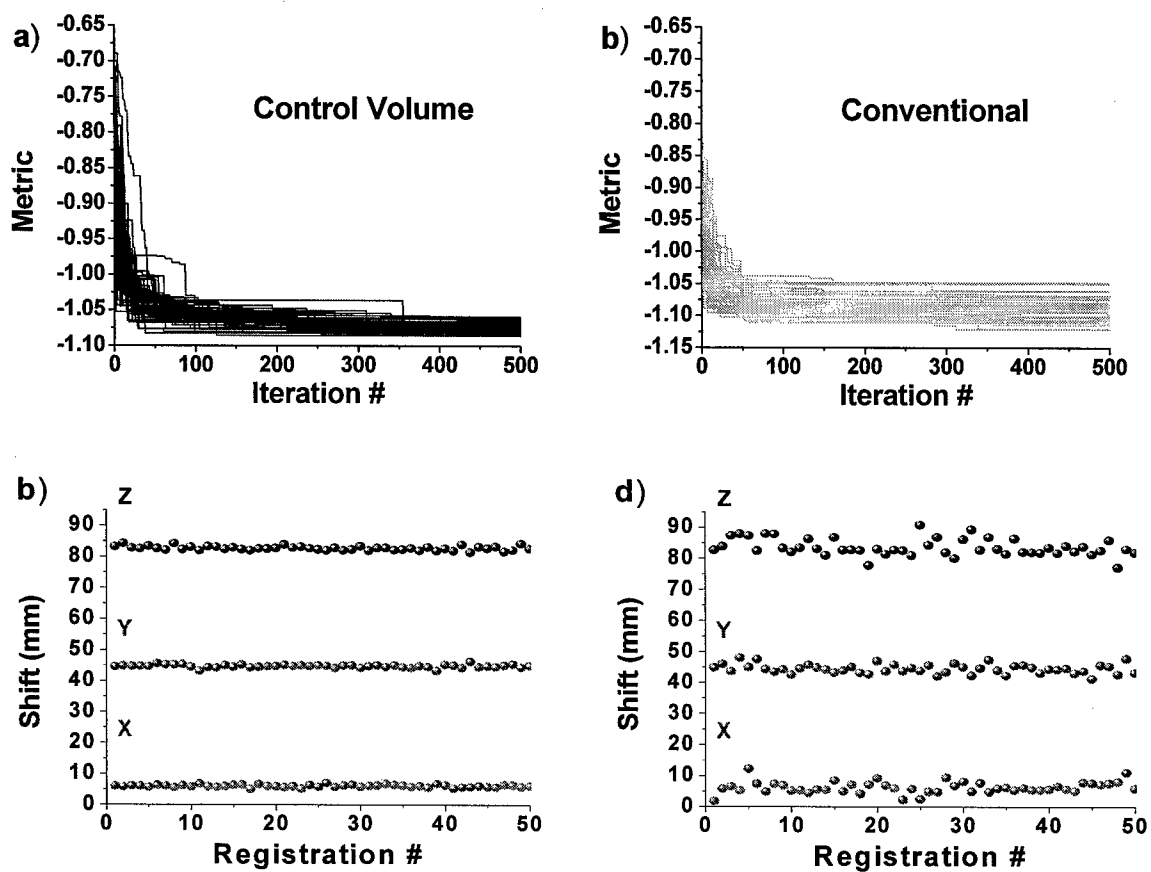
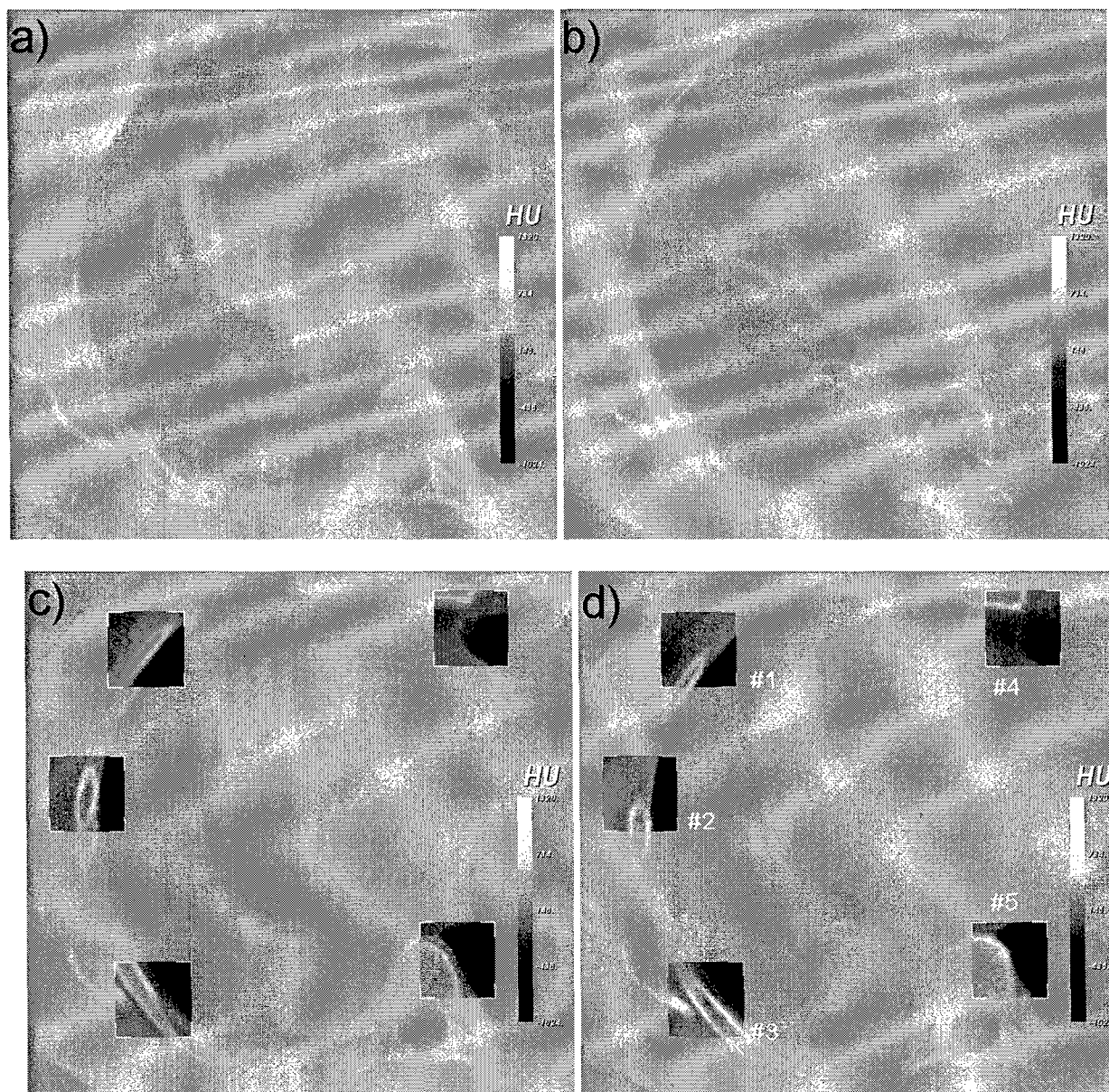
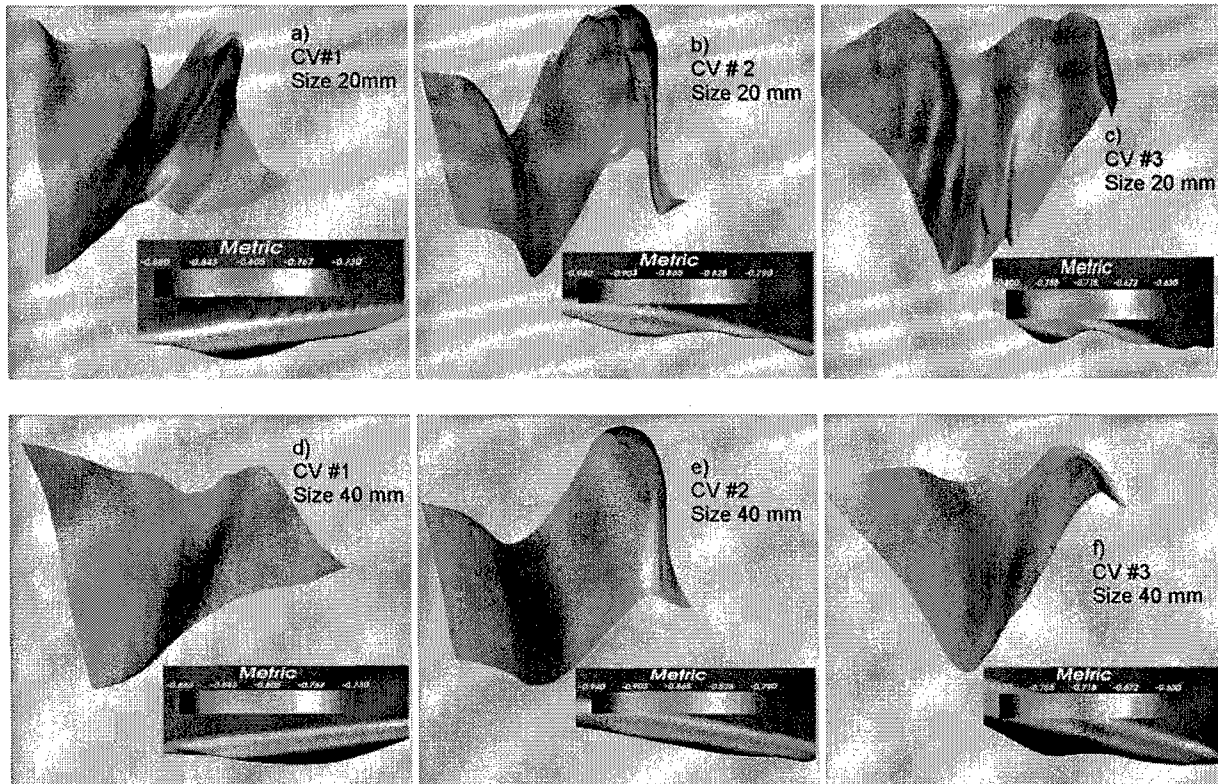


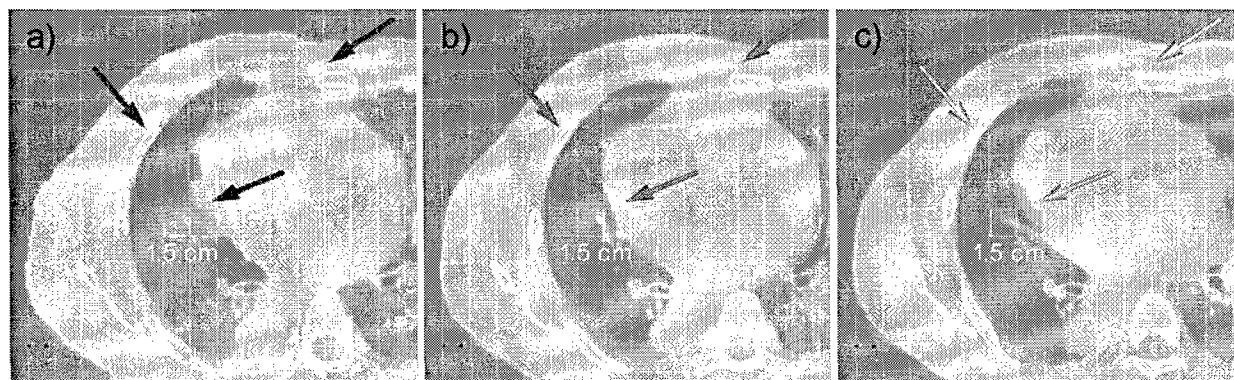
Figure 5

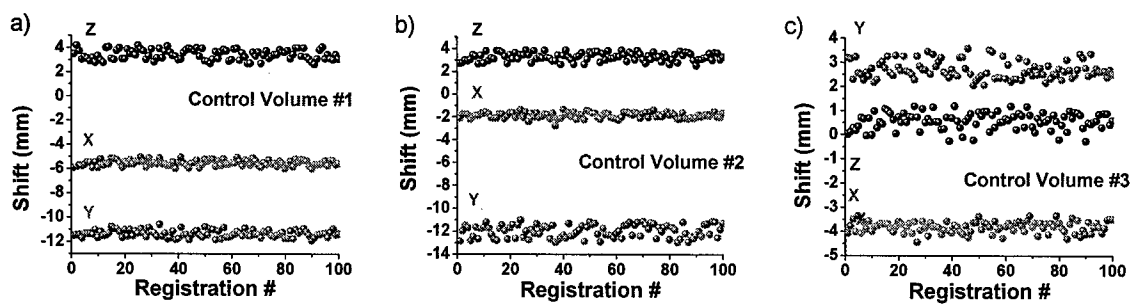












Chapter 2

PHYSICS OF IMRT

LEI XING, PHD, QIUWEN WU, PHD, YONG YANG, PHD, ARTHUR BOYER, PHD

Radiation therapy (RT) as a means of managing cancer has its roots in the discipline of radiology. From the time Roentgen first discovered x-rays, two-dimensional transmission images of the human body provided unprecedented imagery of bony landmarks, allowing radiologists to deduce the location of internal organs. Using planar radiographs, radiologists planned cancer treatments by collimating rectangular fields encompassing the presumed tumor location. Additional blocks placed daily to match marks on the patient's skin and later the use of low-temperature melting dense alloys provided a cookie-cutter approach to treating the two-dimensional projections of the tumor volumes.

Human anatomy and tumor shapes, however, are inherently three-dimensional. By treating a large amount of nearby normal tissue, physicians were limited by the tolerance of the normal tissue they were treating. Additionally, it was not possible to take the three-dimensional structures into consideration because of the limitations of early dose calculations. The advantage of being able to treat a tumor target conformally can be appreciated by a simple example. Assume that the tumor is a sphere of 5 cm in diameter; it would have a volume of 65.4 cc. If one irradiates it with square fields, directed at the six faces of the cube containing the sphere (an anatomic impossibility that we will allow for the sake of making a theoretical point), a high-dose volume would be created within the sphere containing 125 cc. This represents the three-dimensional nonconformal situation. If one were to treat the volume with circular fields, directed toward the sphere from all directions (which, again, is anatomically impossible), the high dose would be limited to the sphere itself. Approximately 60 cc of normal tissue would be spared. The reduction of tissue irradiated is a factor of $6/\pi$ or about half. This reduction in normal tissue irradiation should theoretically improve the therapeutic ratio and allow the tumor target volume to be treated to a higher dose, thereby improving the probability of tumor control. Other factors play critical roles as well. Tumor biology has a great deal to do with the actual tumor control achieved, but the basic idea of reducing normal tis-

sue irradiation is a valid strategy and the goal for managing local tumor control with a minimum of normal tissue complications. The details on radiobiology are discussed in Chapter 3 "Radiobiology of IMRT."

Three-dimensional conformal radiation therapy (3DCRT) is a method of irradiating target volume defined in a three-dimensional anatomic image of the patient with a set of x-ray beams individually shaped to conform the two-dimensional beam's eye view projection of the target. 3DCRT became feasible with the development of computed tomography (CT). The development of spiral and multislice CT scanners has made the acquisition of large data sets practical. The reconstructed images, acquired with patients in the treatment position, provide a model on which geometric and dosimetric computations can be applied. These data sets can be acquired with spiral scanners capable of recording the transmission data needed to reconstruct 50 to 100 transverse image planes spaced 2 to 5 mm apart. Given adequate immobilization devices to help patients achieve and hold their treatment position for the duration of the image acquisition, these fast scanners provide excellent data sets that can be used for treatment planning. The transmission data are used to reconstruct a three-dimensional data set consisting of Hounsfield numbers associated with voxels. The development of the Digital Imaging and Communication in Medicine (DICOM) standard and its various extensions for data exchange has made possible the use of CT data sets acquired with the equipment from one vendor with treatment planning systems from another vendor and the ability to treat patients with equipment from yet another vendor. The transfer of these data over computer networks has improved the efficiency and accuracy of the entire treatment planning and delivery process.

Evolution from 3DCRT to IMRT

Intensity-modulated radiation therapy (IMRT) emerged in clinical practice as a result of the development of 3DCRT

in the 1980s. Although the exact beginning of the modality depends on one's definition of IMRT, it is generally agreed on that the widespread implementation and realization of the technique occurred in the United States in the early 1990s with the commercially available Peacock IMRT planning system and MIMiC fan beam delivery device (North American Scientific, NOMOS Radiation Oncology Division, Cranberry Township, PA).^{1,2} This was then followed by the cone beam multileaf collimator (MLC)-based IMRT in the mid-1990s. MLC allows the rapid and controllable adjustment of field aperture and is thus ideally suited for dynamic radiation beam modulation. In Figure 2-1, different IMRT modalities currently available or under intense investigation are summarized. Physically, a common feature of these IMRT techniques is that they all attempt to enhance control over the three-dimensional dose distribution through the superposition of a large number of independent segmented fields from either a number of fixed directions or from directions distributed on one or multiple arcs.

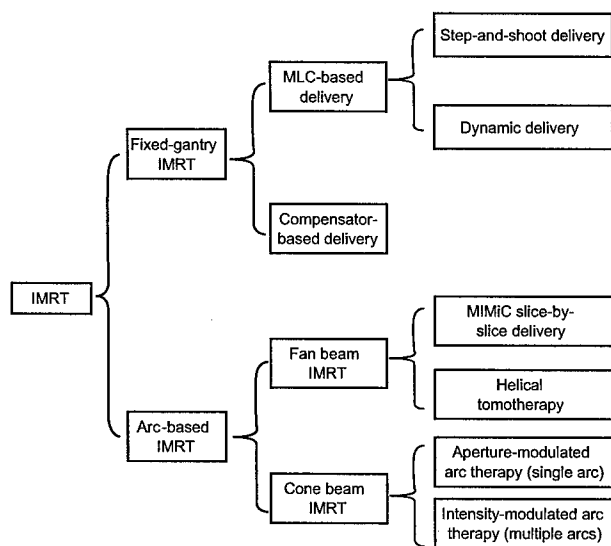
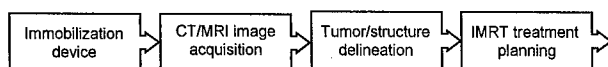


FIGURE 2-1. Currently available intensity-modulated radiation therapy (IMRT) techniques. MLC = multileaf collimator.

Planning



Delivery and verification

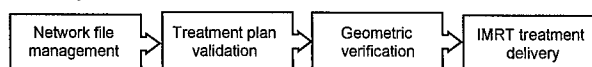


FIGURE 2-2. The intensity-modulated radiation therapy (IMRT) treatment process. CT = computed tomography; MRI = magnetic resonance imaging.

Intensity modulation adds a new degree of freedom to RT planning and provides a more effective means to produce tightly conformal dose distributions in complex treatment situations. The objective of this chapter is to provide an overall comprehension of IMRT and to review the physics aspect of this technology. In Figure 2-2, the overall treatment process of IMRT is illustrated. The key steps involved in the process are discussed in separate sections. In the remainder of this introductory section, we briefly describe the IMRT delivery modes listed in Figure 2-1. Given that fixed-gantry IMRT is by far the most popularly implemented technique, emphasis is given to this mode first.

Fixed-Gantry IMRT

Fixed-gantry IMRT is similar to 3DCRT in that a number of fixed beam directions are used (Figure 2-3A). In this mode, treatment planning is generally done in two steps. First, the dose optimization engine generates a set of intensity profiles, one for each incident beam. Depending on the treatment planning system, the optimized beam profile can be continuous or in a form that is discretized in space and intensity. Without loss of generality, an incident beam is assumed to be already divided into a grid of beamlets, and each beamlet can take a fixed number of intensity levels. The beamlet width (dimension perpendicular to the leaf travel direction) is limited to the MLC leaf width. The beamlet length, or the step size of MLC leaf movement defined as the smallest step in the leaf travel direction, is a parameter specified by the user. A smaller beamlet size or a larger number of intensity levels offers better spatial or intensity

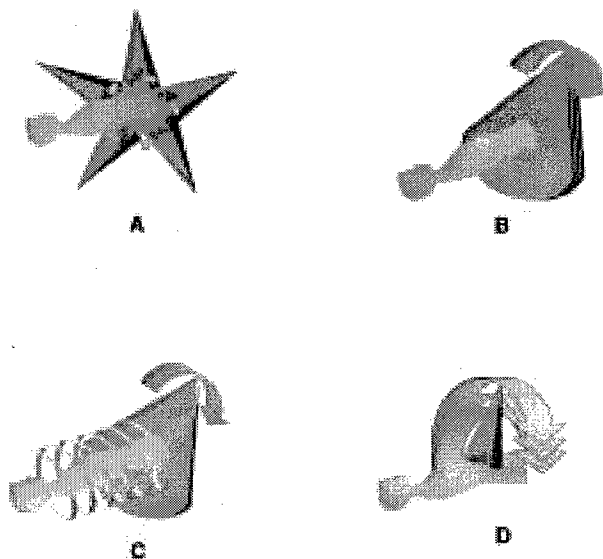


FIGURE 2-3. A schematic drawing of (A) fixed-gantry intensity-modulated radiation therapy (IMRT); (B) slice-by-slice fan beam delivery; (C) tomotherapy delivery; and (D) cone beam-based IMRT.

resolution but requires more MLC segments for delivery. Typically, the size of the beamlet and the number of intensity levels in current IMRT treatment are set to 1×1 cm and 10, respectively. Figure 2-4 shows an example of an intensity map for a head and neck IMRT treatment, obtained using the CORVUS (North American Scientific) inverse planning system. Occasionally, the beamlet size or the number of intensity levels is varied to meet a specific clinical requirement.

There are many ways to produce a desired fluence map. Conceptually, physical compensators are the most straightforward. The most popular delivery technique is, however, based on computer-controlled MLC. In this approach, an intensity map is decomposed into a set of MLC-formed apertures by using a leaf sequencing algorithm. The MLC sequences are recorded in a computer file, which is then used to control the MLC movement for plan delivery. It is important to note that an intensity map, regardless of its shape, can always be expressed as a superposition of a number of segmented fields (for a given intensity map, generally, a number of ways exist for this decomposition, leading to numerous leaf sequencing algorithms). Depending on the relationship between MLC leaf movements and radiation dose delivery, the delivery can generally be divided into step-and-shoot delivery and dynamic modes. The former is the simplest computer-controlled delivery scheme of the fixed-gantry IMRT, in which MLC leaf movements and dose deliveries are done at different instances. A leaf sequence file consists of alternatives of dose-only and motion-only instances. Dynamic delivery differs from a step-and-shoot mode in that leaf movement and dose delivery are realized simultaneously.

Arc-Based IMRT

Arc-based treatment delivery has a long history in RT. An early implementation of this method was the so-called Takahashi arc, in which the beam aperture dynamically follows the beam's eye view projection of the target. Stereotactic radiosurgery based on cylindric cones or micro-MLC often uses the arc delivery technique to "spread" the radiation dose to different regions of the brain to avoid overdosing the normal brain tissue. Conformal arc therapy can produce excellent dose conformation to a simple target. However, the target volumes often exhibit significant deviation from the ideal spherical or ellipsoidal shape. In this case, arc-based IMRT treatment, which was first proposed by Dr. Yu⁴, provides a viable option to improve the dose distributions through intensity modulation. The three different forms of arc-based IMRT deliveries are schematically shown in Figure 2-3, B to D, and their features are summarized below.

Fan Beam IMRT

A schematic drawing of the fan beam IMRT is shown in Figure 2-3B. The delivery is realized on a slice-by-slice manner, in which each slice covers 2 to 4 cm in the longitudinal direction and 20 cm in diameter. North American Scientific's

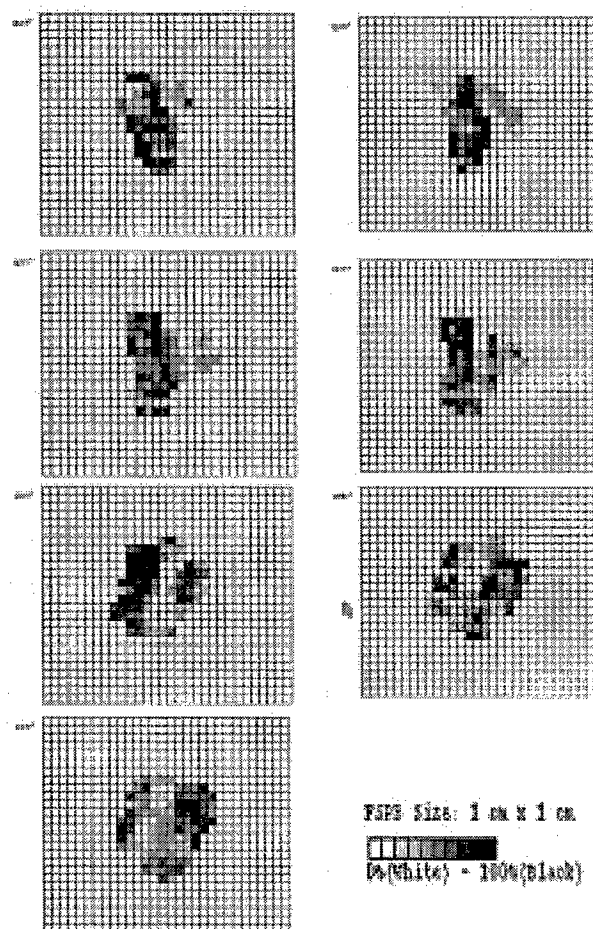


FIGURE 2-4. Intensity patterns of a seven-field intensity-modulated radiation therapy (IMRT) head and neck treatment obtained using the CORVUS IMRT planning system.

Peacock system, which includes the PEACOCK inverse treatment planning system and the MIMiC collimator, is used for this type of treatment. The planning system uses 54 equally spaced beams and optimizes the beamlet maps of each beam. The nominal beamlet sizes on the isocenter plane are 1×0.4 cm, 1×1 cm, and 1×2 cm. An advantage of this modality is that the MIMiC collimator can be retrofitted to an existing linear accelerator without an MLC, allowing IMRT treatment without a substantial hardware upgrade. Use of the arc delivery mode often results in a superior dose distribution in comparison with fixed-gantry IMRT with five to nine beams for deep-seated tumors because of the involvement of a large number of beams in an arc-based treatment.³

Tomotherapy

The tomotherapy machine has recently become commercially available (TomoTherapy Inc., Madison, WI). The delivery is also achieved slice by slice but in a helical (or spiral) fashion in which the couch moves at a constant speed during the gantry rotation (see Figure 2-3C). Radiation from the linear accelerator first passes through a single

set of primary collimator jaws, which shape the beam into a rectangular slit that is 40 cm long and up to 5 cm wide at the isocenter. The MLC that is used to modulate the beam intensity consists of 64 tungsten leaves that move across a narrow opening to control the radiation passing through to the target. The computer-controlled MLC has two sets of interlaced leaves that move in and out very rapidly to constantly modulate the beam.

Cone Beam–Based IMRT

To date, the majority of work on arc-based IMRT has been focused on modulated fan beams, and little development has been done using cone beams. The concept of intensity-modulated arc therapy (IMAT) was first proposed in 1995 (see Figure 2-3D), and manufacturers have provided the technical capability for dynamic arc delivery.⁴ However, IMAT has not been widely implemented. The lack of enthusiasm for IMAT stems in part from the shortage of effective planning tools and reliable quality assurance (QA) procedures. Reports from several institutions, however, support the notion that a cone beam–based arc technique can generate superior dose distributions, at least for some deep-seated tumors.^{5–8}

Cone beam arcs use the arc feature of fan beam IMRT yet take advantage of the cone beam modulation of the fixed-gantry IMRT. To compute dose distributions, an arc is approximated by many fixed fields at small intervals of gantry rotation. Physically, however, the achievement of intensity modulation for cone beam delivery is less straightforward in comparison with its fan beam counterpart. Unlike a slice-by-slice delivery, in which the radiation across the slice can be segmentally blocked from the side by multiple independent vanes, the MLC-shaped aperture cannot change from one shape to another fast enough as the gantry rotates. This problem can be solved, at least in principle, by lowering the gantry rotation speed because, in reality, it is the relative speed between the gantry rotation and MLC leaf movement that determines the level of achievable intensity modulation.

An alternative approach is to use multiple cone beam arcs, as proposed by Yu.⁴ At each gantry angle, the beam is considered to be a superposition of a series of subfields, each with uniform intensity from these arcs. When a single arc is used for treatment, the technique is sometimes called aperture-modulated arc therapy. At this time, there are no studies defining how many arcs are sufficient for any disease sites. In Figure 2-5, a comparison of average dose-volume histograms (DVHs) of 3DCRT, IMRT, and IMAT prostate plans for ten patients with prostate cancer is shown.⁹ The solid line is IMAT, the dotted line is IMRT, and the dot-dash line is 3DCRT. It is evident that IMAT yields better target coverage and improved bladder and rectum sparing in comparison with fixed-gantry IMRT. Finally, being able to modulate the dose rate while the gantry rotates is a desirable feature, further enhancing the performance of the cone beam arc–based IMRT. To date, however, no linear accelerator manufacturers have provided such technical capability in the clinical mode.

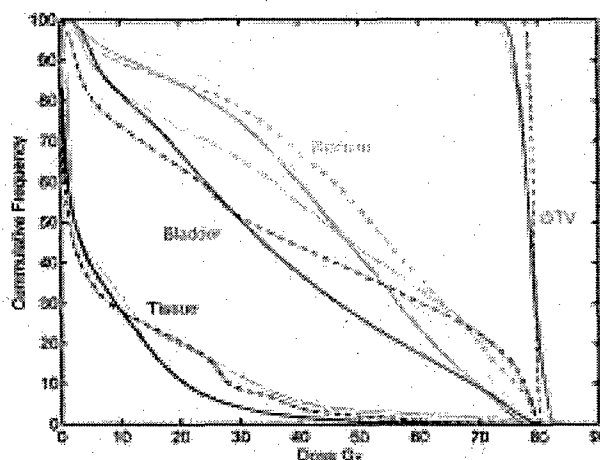


FIGURE 2-5. Average dose-volume histograms for 10 patients with prostate cancer planned using three-dimensional conformal radiation therapy (*dot-dash line*), intensity-modulated radiation therapy (*dotted line*), and intensity-modulated arc therapy (*solid line*). GTV = gross tumor volume. (To view a color version of this image, please refer to the CD-ROM.)

Treatment Planning

RT planning requires the calculation of a set of parameters for the delivery of a radiation dose to the patient. Although manual forward planning may be possible in some simple cases (see the examples below), computer optimization of the beam parameters is almost always used for IMRT treatment planning because of the vast size of search space involved in the problem. In general, this is realized using an inverse treatment planning technique, which derives the optimal beam parameters by starting from a prescribed or desired dose distribution. Although the details of the inverse planning calculation depend on the delivery method, the principle behind the algorithms is essentially the same. Inverse treatment planning is, in fact, a special case of general inverse problem encountered in the sciences and engineering, which attempt to derive the optimal input parameters that will produce the desired output. Before discussing the inverse planning algorithms in detail, it is illustrative to briefly summarize the features of the forward planning approach.

Forward Planning for Segment-Based Treatment

There are two aspects in RT planning: dose conformity and dose uniformity inside the target. What it takes to accomplish the two goals may be different. When the shape of the target is regular and/or when only two or three incident beams are employed, the isodose shaping can often be achieved by beam shaping with an MLC. To achieve a uniform dose distribution within the target volume, one only needs to accommodate the geometric variation of the external contour. Physical

or dynamic wedges are usually used if the patient contour changes monotonically or in some simple hinge field arrangements. In a more general situation, additional MLC-shaped field segments can be introduced to boost a "cold" region or reduce a "hot" region. Examples of this type of case include but are not limited to opposed tangent field breast treatment and anterior-posterior treatment of Hodgkin's disease. For illustration purpose, a forward multisegment breast treatment plan is considered.

The multisegment breast plan starts with the standard opposed tangent fields. In many cases of breast cancer, obtaining a uniform dose within the target volume could be problematic when this approach used. To improve on this, one may proceed to sequentially introduce additional MLC field segments to one or both beam directions to boost the cold region(s) under the guidance of dose distribution in the plane perpendicular to the incident beam direction. Figure 2-6 illustrates the three segments of the lateral and



FIGURE 2-6. Opposed tangential fields for the treatment of a patient with left-sided breast cancer (*top row*). The middle and bottom rows are the multileaf collimator shapes of the three segments of the medial and lateral fields chosen to improve dose uniformity within the treatment volume. (To view a color version of this image, please refer to the CD-ROM.)

medial fields. In this plan, the first segments in the lateral and medial fields and their relative weights are determined using conventional techniques. A physical wedge of 30° is placed on the lateral beam. The two additional segments in each beam direction are then introduced sequentially, and their weights and apertures are adjusted using trial and error to achieve a more uniform dose distribution. The isodose distributions for both plans are shown in Figure 2-7. The maximum dose and the volume receiving a high dose in a multisegment plan are significantly reduced.

Multisegment-based forward planning techniques can be applied only to some relatively simple cases in which the high-dose region is primarily defined by the conventional treatment fields. When isodose conformity to an irregularly shaped target is needed, multiple beams (typically more than five) with a higher level of intensity modulation are needed. In this situation, it becomes tedious to use forward planning-based approaches, and more sophisticated inverse planning techniques become necessary.

Inverse Planning

Inverse planning uses a computer optimization algorithm to determine the optimal beam parameters that lead to a solution as close as possible to the desired output. Mathematically, the problems of image reconstruction, image restoration, signal process, and investment portfolio

management can all be formulated as an inverse problem. Roughly speaking, inverse problems can be described as problems in which the output or consequences are known but not the cause. The difference between various treatment planning systems lies in the specifications of the input and output parameters and the criteria used to select the final solution. Specific to RT, the output is generally specified by a desired dose distribution, a set of desired DVHs, or even the tumor control probability (TCP) and normal tissue complication probability (NTCP) for the involved structures. The input parameters to be optimized depend on the delivery scheme. Typically, the number of beams and their incident directions are determined empirically before dose optimization. Each incident beam is discretized into a bixel map (the bixel or beamlet size is typically $1 \times 1 \text{ cm}^2$). The task of inverse planning is then to determine the optimal bixel map or the relative weights of all of the beamlets.

To better appreciate the problem, assume that six incident beams are used for an IMRT treatment. If each beam is divided into 100 beamlets and each beamlet has 10 permissible intensity levels, there would be $10^6 \times 100$ physically realizable plans. It can be shown that the number of physically realizable solutions for a six-field 3DCRT plan is much less than this number. When wedges are not used, there are 10^6 physically feasible solutions (many of these can be immediately eliminated from being a candidate treatment plan because they do not produce clinically acceptable dose distributions). For a given desired dose distribution D_0 , the task is to find a solution D in the physically feasible solution pool $\{D\}$ that is the same as D_0 or, more appropriately, differs the least from D_0 . There are many ways to pick a D that is a "good" representation of the prescribed dose D_0 . A commonly used approach for plan optimization is to minimize the distance between D and D_0 in the L^2 norm. For therapeutic applications, it is common to introduce an importance factor r_σ to control the relative importance of the structure σ . This leads to the following quadratic objective function:

$$F = \frac{1}{N} \sum_{\sigma} r_{\sigma} [D_{\sigma}(n) - D_0(n)]^2, \quad (1)$$

where r_σ is the importance factor that weights the importance of the structure σ and parameterizes our clinical trade-off strategy and D_0 and D_n are prescribed and calculated doses, respectively. Optimization of this function is essentially a least squares type of estimation in statistical analysis. In addition to equation 1, many other types of objective functions have been proposed for plan optimization. The construct of the objective function plays a crucial role for the success of IMRT treatment and is worthy of detailed discussion.

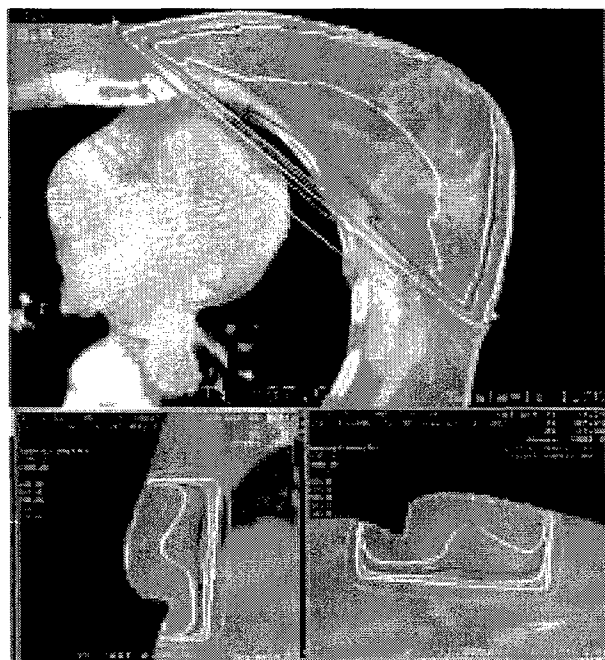


FIGURE 2-7. Conventional opposed tangential breast (*left*) and multi-segment (*right*) plans. The isodose lines are (from inside to outside) 105, 100, 95, 90, 80, 50, 20, and 10%. A dose of 50.4 Gy was prescribed to the 90% line. Note that the use of multiple segments improves the dose uniformity to the breast. (To view a color version of this image, please refer to the CD-ROM.)

Models and Model Parameters of Inverse Planning

A common feature of all inverse problems is that they are generally underdetermined and ill-posed. The selection of the final solution depends on the underlying assumption of the model. The objective function quantitatively ranks a candidate treatment plan, and the optimization of the function yields the optimal parameters. In conventional treatment planning, the objective function depends on beam weights, wedge angles, and orientations, whereas in IMRT, it is a function of the beamlet weights. Ideally, an objective function would mimic the decision-making of experienced oncologists and planners. It would rank a given solution (corresponding to a set of parameters) in a way consistent with clinical judgment. In practice, however, a gap exists between mathematical modeling and clinical decision-making, and much effort is being devoted to derive clinically meaningful objective functions for inverse planning. Because the optimization results depend strongly on the objective function, there is inevitably subjectivity associated with the various dose optimization schemes. Therefore, it is essential for physicians to carefully evaluate a treatment plan after optimization to ensure that the "optimal" solution makes clinical sense. Otherwise, the success of an optimization is, at best, mathematical.

If an optimization algorithm is to have a genuine impact on clinical practice, it should incorporate all of the dosimetric and radiobiologic knowledge plus an algorithm for modeling the way in which radiation oncologists and patients balance the risks and benefits. Despite the availability of high-speed computers, state-of-the-art inverse planning algorithms, and improved imaging modalities, we are still a long way from generating truly optimized IMRT treatment plans. For convenience, it is appropriate to classify the currently available dose optimization methods into four categories: (1) dose based, (2) clinical knowledge based, (3) equivalent uniform dose (EUD) based, and (4) TCP or NTCP based. The underlying difference between these models lies in which end points are used to evaluate the treatment plan or which fundamental quantities are used to define the optimal plan. In reality, each type of inverse planning formalism has its own pros and cons in coping with the clinical decision-making process and in practical implementation. These are briefly summarized below.

Dose-Based Formalism

The dose and/or dose volume-based optimization is concerned with accurate dose distributions or DVHs of the involved structures. The quadratic objective function given in equation 1 represents an example of this type. Frequently, DVHs and other physical constraints are imposed to describe certain clinical requirements. The dose or dose volume prescriptions are used implicitly as surrogates of the desired clinical outcome. At this point, the dose-based approach is the most widely employed method, as is evidenced by the

fact that all commercial IMRT planning systems have chosen dose-based ranking as the starting point. There are several reasons for this. First, the physical dose objectives reflect the majority of the clinical practice. Although biologic models are available in both research and clinical systems, the uncertainty associated with the predictions often outweighs their guidance. Dose-based objectives will remain the dominant modality of optimization and evaluation for some time. Second, the physical dose is closely related to the optimization parameters, and simple mathematical models, such as the quadratic dose difference expressed in equation 1, can be effectively used.

Clinical Knowledge-Based Formalism

It is highly desirable to incorporate clinical end points in guiding the treatment plan optimization process. The currently available dose-based objective functions do not truly reflect the nonlinear relationship between dose and the response of tumors and normal tissues. In reality, the dose dependence of the clinical end point of a structure may be degenerate in the sense that a given clinical end point may be caused by a variety of dose distributions or DVHs. For the parotid glands, for instance, it is known that the clinical end point is the same if 15 Gy is delivered to 67% of the volume, if 30 Gy is delivered to 45% of the volume, or if 45 Gy is delivered to 24% of the volume. If the dose-based objective function, equation 1, is used, the rankings for the three different scenarios would be different. Even with the use of dose-volume constraints, it is difficult, if not impossible, to incorporate this type of knowledge to correctly model the behavior of the organ in response to radiation. Indeed, a constraint in optimization acts as a "boundary condition" during the optimization (there are methods of treating constrained optimization problem into an equivalent unconstrained one, with a different objective function) and does not change the rankings of dosimetrically different plans.

To overcome these dilemmas, a clinical knowledge-based optimization scheme has recently been developed by Yang and Xing.¹⁰ The central theme of the approach is that clinical outcome data should be used to direct the plan optimization process. In this approach, the quality of a treatment plan is measured by a heuristically constructed objective function which depends not only on the dosimetric properties but also the dose-volume status, which makes it possible to take advantage of the existing outcome data of the involved organs. For the parotid glands, for instance, the three different DVHs mentioned above will be scored equally by the objective function. The final dose distribution or DVHs of the glands will be determined by the optimization algorithm with the consideration of the requirements of other structures. If one of the three possibilities needs to be selected, the one that yields better scores in other involved structures will be favored by the algorithm. The specifics of the plan selection process will, of course, depend on the geometric and dosimetric details of the particular patient.

It is important to emphasize that, at this point, clinical outcome data are sparse and underdetermined and may have large uncertainties. By “underdetermined,” we mean that there are not enough clinical data points available to objectively rank all realizable plans. Thus, it is necessary to produce an interpolation/extrapolation scheme for plan ranking. A sensible approach has also been provided in Yang and Xing’s work based on the well-known dose response model.¹⁰ The clinical knowledge-based model allows one to more objectively rank treatment plans according to their clinical merits without relying on biological index-based or EUD-based prescriptions.

EUD-Based Formalism

Optimization of the dose distributions can also be cast into the realm of EUD, which is one level higher in terms of the use of biologic information.^{11,12} The EUD is defined as the biologically equivalent dose, which, if given uniformly, leads to the same cell kill as the actual nonuniform dose distribution. It can be expressed as follows:

$$EUD = \left(\frac{1}{N} \sum_i D_i^a \right)^{\frac{1}{a}} \quad (2)$$

In this expression, N is the number of voxels in the anatomic structure of interest, D_i is the dose in the i th voxel, and a is the tumor- or normal tissue-specific parameter that describes the dose-volume effect. This formulation of EUD is based on the power law dependence of the response of a complex biologic system to a stimulus.

EUD exhibits a dose-response relationship similar to that of the traditional biologic indices. Therefore, it can be a surrogate for them and, in the meantime, is closely related to the physical dose. The objective function based on EUD can be expressed in the following:

$$F = \prod_j f_j \quad (3)$$

where the component subscore f_j may be either

$$f_T = \frac{1}{1 + \left(\frac{EUD_0}{EUD} \right)^n}$$

for tumors or

$$f_{OAR} = \frac{1}{1 + \left(\frac{EUD}{EUD_0} \right)^n}$$

for normal tissues. There are several advantages of EUD-based optimization approaches: (1) the formulae are simple, (2) the formulae can be applied to both tumors and organs at risk (OAR) using different parameters, and (3) there are fewer planning parameters than dose volume-based or other biologic indices-based optimization. It has been shown that EUD-based optimization can provide the same or better coverage of targets as dose volume-based optimization and that it offers significantly better protection of OAR. These improvements in the dose distributions to OAR may be due to the fact that there is a larger search space available in EUD optimization because the constraint, or the objective, is determined on the basis of the whole organ rather than the partial volume of the structure. Thus, EUD optimization can be used to search for and evaluate multiple plans that may have different DVHs but the same EUDs. Figure 2-8 shows the dose distributions for IMRT plans optimized using dose volume-based and EUD-based objective functions for a patient with prostate cancer. The OAR are the rectum and bladder. All plans used identical configurations of five coplanar 18 MV photon beams placed at equally spaced gantry angles. The plans were normalized to deliver the prescription dose of 70 Gy to 99% of the target volume. It is clear that, for the same minimal target dose, sparing of the OAR is greatly improved in the EUD-based plan. Furthermore, a sharp dose gradient at the interface between the target and OAR is realized.

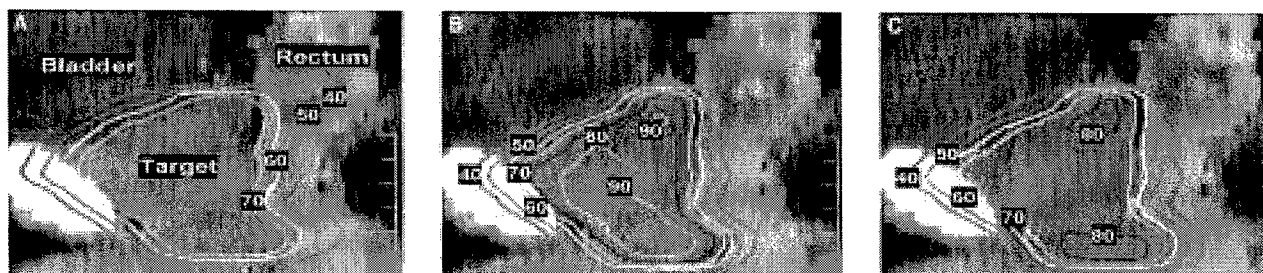


FIGURE 2-8. Sagittal isodose distributions for prostate intensity-modulated radiation therapy plans designed using (A) dose volume-based criteria; (B) equivalent uniform dose (EUD)-based criteria; and (C) EUD-based criteria with target inhomogeneity constraints. Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

The EUD concept can also be incorporated into the framework of physical dose optimization, such as the method of projection over the convex set.¹³ In this method, EUD is implemented as an optimization constraint. At each iteration of the optimization, if an organ violates an EUD constraint, a new dose distribution is calculated by projecting the current one onto the convex set of all dose distributions fulfilling the EUD constraint. The cost is slightly more iterations than pure physically constrained optimization. This algorithm is easy to implement and provides better dose sparing of parallel structure organs for which physical constraints may be difficult to define.

Biologic Model-Based Formalism

Biologic model-based optimization proponents argue that plan optimization should be guided by estimates of biologic effects. The biologic effect and the radiation parameters are linked by the radiation dose through the use of a dose-response function. The relationship between the two is not, however, a one-to-one correspondence. A given biologic end point may be produced by many possible dose distributions, which would generally not be equally scored if a dose-based model was used. In principle, biologically based models are most relevant for RT plan ranking.^{11,14-21} However, the dose-response function of various structures is not sufficiently understood, and at this point, there is considerable controversy about the models for computing dose-response indices and their use in optimization.

The treatment objective in biologic model-based inverse planning is usually stated as the maximization of the TCP while maintaining the NTCP to within acceptable levels.^{22,23} Physical constraints on dose and dose volume are often introduced to ensure that the results are consistent with the clinical judgment of the physicians. Brahme and Kallman and colleagues used the probability of uncomplicated control, P_+ , in their formalism.^{14,24} Practically, the use of dose-response indices for optimization might also pose some problems. For instance, dose response-based optimization may lead to very inhomogeneous target dose distributions. Furthermore, it is difficult for clinicians to specify the optimization criteria in terms of certain dose-response indices (eg, TCP, NTCP, and P_+). This difficulty becomes even more significant when two or more independently optimized plans are to be combined because it is impractical to specify the desired TCP and NTCP of the component plans. Because of these problems, the use of biologic model-based dose optimization has mainly been limited to the research setting and little effort has been made to implement these into commercial IMRT planning systems.

Model Parameters

Any dose optimization framework must deal with trade-offs between the target and OAR.²⁵ Generally, the objectives of different structures are multifaceted and incommensurable. A combination of the objectives is usu-

ally done to form a single objective function. In this process, a set of importance factors is often incorporated into the objective function to parameterize trade-off strategies and prioritize the dose conformity in different anatomic structures. Whereas the general formalism remains the same, different sets of importance factors characterize plans of obviously different flavor and thus determine the final plan. One of the major difficulties is that the influence of these weighting factors on the final solutions is not known until the dose optimization is done, necessitating a trial-and-error determination of the parameters. In most (if not all) of the currently available planning systems, the values of the weighting factors are presented to the user as optimization parameters. A good understanding of the role of these parameters and suitable training on how to empirically determine the parameters are required.

It is possible to use an iterative algorithm to estimate the weighting factors numerically.²⁵ Plan selection is done in two steps. First, a set of importance factors is chosen, and the beam profiles are optimized under the guidance of a quadratic objective function using an iterative algorithm. The "optimal" plan is then evaluated by a decision function, in which the corresponding trade-off parameters are more easily determinable based on some simple considerations.²⁵ The importance factors in the objective function are adjusted iteratively toward the direction of improving the ranking of the plan. For every change in the importance factors, the beam parameters are reoptimized. Even though further refinement of the plan may still be needed in selected cases, the technique provides a good starting point for planning.

Dose Optimization Algorithms

Although the modeling of RT treatment is of paramount importance, the optimization of the selected multidimensional objective function provides a vehicle to obtain the optimal solutions. The task of an optimization algorithm is to find the combination of beam parameters that optimize the chosen objective function, possibly subject to some constraints. Numerous algorithms have been developed for the optimization of a multidimensional function in the sciences and engineering over the years, and there is a vast literature on the subject. Generally speaking, the selection of an optimization technique depends on the specific form of the objective function and the imposed constraints. In practice, even for the same class of problem, more than one algorithm may exist for achieving the same goal, and the detailed implementation of different algorithms can be quite different. Many optimization techniques have been used for RT inverse planning. Here we briefly describe a few approaches to illustrate how a multiobjective objective function is optimized and the pros and cons of these common techniques (see Chapter 10 "Treatment Planning").

Iterative Algorithms

The iterative method is perhaps the most widely implemented technique in RT optimization. Starting with an initial approximate solution, it generates a sequence of solutions that converge on the optimal one. For large systems, especially large linear systems, iterative methods prove to be efficient in terms of computer storage and computational time. The available iterative techniques can generally be grouped as non-derivative-based and derivative-based methods. The former incorporates only an objective function value calculation with some systematic method to search the solution space. This technique is generally intuitive, easy to implement, and particularly suitable for simple systems and educational illustration. For a complex system, the convergence behavior may not be as good as more sophisticated gradient-based search techniques. The computational cost and poor convergence in this situation may outweigh the benefit of avoiding derivative calculations.

As an example, Figure 2-9 illustrates the flowchart of an algebraic iterative inverse planning technique (AIIPT) described by Xing and colleagues.^{26,27} A schematic drawing of calculation pixels and bins in the AIIPT calculation is shown in Figure 2-10. The algorithm was generalized from the algebraic reconstruction technique (ART) based on the analogy between rotational RT optimization and tomographic image reconstruction. In the AIIPT algorithm, voxels are examined in sequence, and corrections are made immediately after a pixel is addressed. The successive treatment of the system eventually leads to an optimized solution. A geometric interpretation of ART has been published.²⁸

Algebraic iterative method:

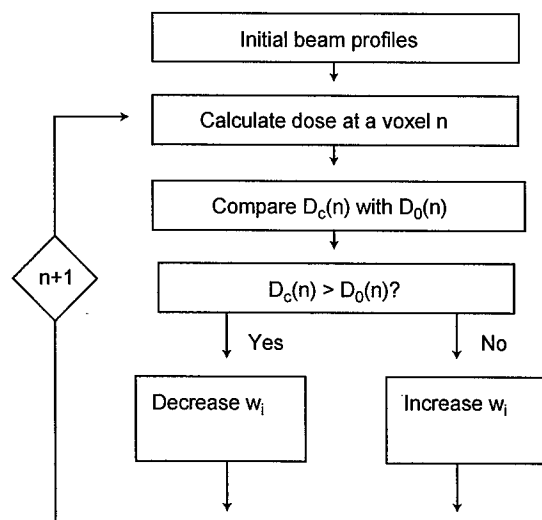


FIGURE 2-9. Flowchart of the algebraic iterative inverse planning technique.

The iterative process is described by the following operations: (1) assume an initial set of beam profiles; (2) compute the dose at a voxel; (3) compare the calculated and prescribed doses; (4) obtain correction factors to the beamlets that irradiate the voxel; and (5) apply the corrections to the contributing beamlets and then repeat from step 2 for the next voxel (go back to the first voxel and increase the iteration index by one after all voxels are addressed). This process is repeated until the desired accuracy is achieved. The simultaneous iterative inverse planning and least squares inverse treatment planning algorithms also fall into the same category of the nonderivative method.^{26,27} A similar algorithm with a multiplicative beamlet updating scheme was described by Jones and Hoban.²⁹

Various gradient-based methods have been successfully applied to RT plan optimization and implemented in commercial IMRT planning systems. A general class of iterative algorithms can be written as

$$\mathbf{I}^{new} = \mathbf{I}^{old} - \lambda \mathbf{M}^{old} \nabla F(\mathbf{I}^{old}) \quad (4)$$

where \mathbf{I} is the fluence vector, \mathbf{M}^{old} is a matrix, and λ is a positive parameter. When \mathbf{M}^{old} is a unit matrix, equation 5 is the well-known steepest descent algorithm, whereas when $\mathbf{M}^{old} = \mathbf{H}^{old}$, it describes Newton's method, in which \mathbf{H}^{old} is the inverse of the Hessian matrix.³⁰ In the steepest descent algorithm, from a set of fluence functions, \mathbf{I}^{old} , we search along the direction of the negative gradient, $-\nabla F(\mathbf{I}^{old})$, to a minimum on this line; this minimum is taken to be \mathbf{I}^{new} .

It is fair to say that the choice of a specific algorithm

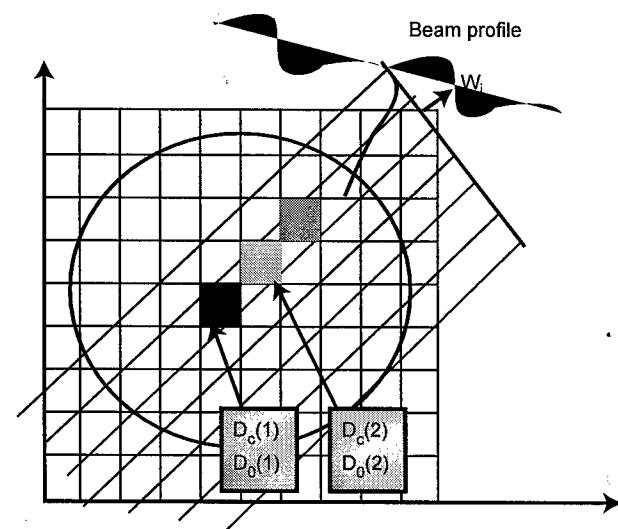


FIGURE 2-10. Pixel and bin configurations used in the algebraic iterative inverse planning technique. In this figure, w_i is the beamlet weight and D_c and D_0 represent the calculated and prescribed doses, respectively. Reproduced with permission.

to solve the inverse planning problem is not unique and is determined by the problem at hand and, to a certain extent, by personal preference. Generally, the iterative approach works well for a nonconvex objective function, and the solution can be trapped in local minima for systems with a complicated form of the objective function. Several commercial systems provide both iterative and stochastic optimizers so that users have the tool to compare the functionality of different approaches and, more importantly, to independently check an optimization calculation.

Computer-Simulated Annealing

The simulated annealing method³¹ is an extension of the original Monte Carlo simulation algorithm introduced by Metropolis and colleagues.³² It attempts to find the optimal solution by mimicking the behavior of a system of interacting particles that are progressively cooled and allowed to maintain thermal equilibrium while reaching the ground state. In physical annealing, the system is heated, thereby conferring randomness to each component. As a result, each variable can temporarily assume a value that is energetically unfavorable, and the system explores configurations that have a higher energy. The fundamental principle here is that even at moderately high temperatures, the system slightly favors regions in the configuration space that are overall lower in energy and hence more likely to contain the global minimum. The algorithm employs a random search that not only accepts changes that decrease objective function but also some changes that increase it. The probability for accepting a trial configuration is controlled by the temperature and is given by

$$P = \begin{cases} 1 & \text{If } \Delta F < 0 \\ \exp(-\frac{\Delta F}{T}) & \text{Otherwise} \end{cases} \quad (5)$$

where ΔF is the increase of the objective function and T is the system temperature. The temperature is gradually lowered according to an empirically chosen cooling schedule.^{33,34} As the temperature is slowly reduced, the probability of accepting a trial configuration with a higher objective function value is reduced. The starting temperature is chosen to be higher than the largest value of objective function calculated for a random set of variable configurations. In principle, this algorithm is capable of finding the global minimum of a multidimensional objective function even when local minima exist. For more details about the simulated annealing algorithm, readers are referred elsewhere.³⁴⁻³⁶

Other Optimization Algorithms

In addition to the iterative and simulated annealing algorithms, many other types of optimization approaches have been employed for therapeutic plan optimization. Linear

programming was applied to the dose optimization of 3DCRT plans and cyberknife plans. The utility of filtered backprojection from CT image reconstruction has also been explored by several researchers.³⁷⁻³⁹

The constrained least square algorithm⁴⁰ was employed to optimize 3DCRT plans^{41,42} and IMRT plans. Constrained optimization of a linear system can be viewed in two ways. One involves transformation of the problem into a reduced space. Another approach is to work with the lagrangian function and to obtain the solution of the system by a direct matrix manipulation. In this way, a priori knowledge of the variance of the system variables can be included as a lagrangian multiplier. Without repeatedly invoking the dose calculation, this algorithm allows one to obtain the optimal solution of the system with significantly increased computational speed, providing a fast interactive planning environment for IMRT planning.

Mixed integer programming technique was used to generate treatment plans for linear accelerator-based radio-surgery,⁴³ IMRT,⁴⁴ and 3DCRT.⁴⁵ Lee and Zaider also applied integer programming for permanent prostate implant planning.⁴⁶ The mixed integer programming models incorporate strict dose restrictions on the tumor volume and constraints on the desired number of beams, isocenters, couch angles, and gantry angles. The goal is to deliver the full prescription dose uniformly to the tumor volume while minimizing excess radiation to the surrounding normal tissue. Hou and colleagues used simulated dynamics in a classic system of interacting particles for IMRT optimization.⁴⁷ In this approach, an analogy is established between intensity profile optimization in IMRT and relaxation to the equilibrium configuration in a dynamic system. Dose-volume constraints are handled by placing hard constraints on partial volumes. The genetic algorithm is another widely used approach in sciences and engineering and has found some preliminary application in RT dose optimization.⁴⁸⁻⁵¹

For all of their complexity, the algorithms to optimize a multidimensional function are routine mathematical procedures. In general, simulated annealing and genetic algorithms are powerful approaches, but excessive computation time is a drawback to their clinical application. Treatment planning based on filtered backprojection and direct Fourier transformation have difficulty handling the negative fluence problem and are not generally applicable for an arbitrary dose prescription and kernel. Iterative methods are widely used to optimize a multidimensional objective function by starting with an initial approximate solution and generating a sequence of solutions that converge to the optimal solution of the system.

It is useful to note that much effort has also been devoted to formulate the problem into a more effective mathematical framework. For example, Xing and Lian and their colleagues introduced a new concept of a preference function and recast the problem into the framework of Bayesian

statistical analysis.^{35,52-54} In this approach, instead of a rigid prescription dose, a range of prescription doses prioritized by the preference function is allowed. The rationale here is that since a rigid prescription is not achievable and the final solution will deviate from it anyway, we would have much better chance to obtain what we want if we could inform the system with some a priori information about our preferences on different possible scenarios (instead of leaving the decision-making totally to the computer). The techniques developed over the years in statistical decision-making can be easily extended to RT plan optimization problem. The primary advantage of the technique is that it enables one to effectively incorporate the existing clinical knowledge or other prior knowledge into inverse planning. When the prescribed dose takes a single rigid value, the above formula becomes identical to the conventional least squares approach or alike. Maximum likelihood estimation^{55,56} or the maximum entropy approach⁵⁷ also represents a special case of the formalism. Finally, it is interesting to point out that various techniques in related fields such as neural networks⁵⁸ and fuzzy logic⁵⁹ are also being translated for RT dose optimization.

Practical Aspects of IMRT Planning

Inverse planning is a computer-based decision-making technique that derives the optimal treatment plan by starting with a set of desired doses or DVHs prescribed to the target and normal tissues. To use an inverse planning system to generate a treatment plan, one must delineate the tumor volume and sensitive structures, for which dose avoidances are desired. This differs from conventional planning, in which the target volume is often defined directly on the portal films (see Chapter 11 "Plan Evaluation"). If target contours need to be altered after a conventional treatment plan is obtained or during a course of treatment, it is usually achieved by modifying the positions of the corresponding MLC (or by modifying a block). In inverse treatment planning, however, the beam profiles and beam apertures are derived by the system, and any change in the target volume requires reoptimization of the plan. Moreover, all of the tasks following IMRT planning, such as patient-specific QA and data entry, need to be repeated.

IMRT planning is still inherently a trial-and-error process owing to the large number of input parameters.⁶⁰ The trial-and-error process here is quite different from that in 3DCRT, in which intuition and previous experience can be easily used to guide the planning process. In an anterior-posterior treatment, for example, if the dose in the anterior region is higher than that of the posterior region, one can simply increase the weight of the posterior field. This type of guidance is lost in inverse planning, and, frequently, the trial-and-error process has to proceed in a "blind-guessing" fashion because the influence of most of the system parameters is not known until the dose optimization is complete. A good understanding of the effect of treatment planning parameters used

in optimization on the resultant dose distribution is necessary to carry out the planning and the plan "tweaking" process. Recently, tools for assisting the interactive planning have emerged. The dose shaping technique described below is one example. Hopefully, this type of research will make clinical inverse planning more straightforward in the future.

Plan review is an important aspect of IMRT. In inverse planning, an objective function is constructed based on general physical, dosimetric, and biologic considerations and is defined as a global quantity.⁵² The translation of the treatment objectives to a single objective function is at best an approximation. Just like any data reduction or compression scheme, there is a loss of information with regard to the characteristic of the individual data point. Even with the best possible objective function, the optimal solution may still not represent the best clinical solution in every aspect. It is important to review the plan to ensure that the final solution is consistent with clinical judgment. IMRT plan evaluation tools vary from one commercial planning system to another. Typically, they include isodose distributions in axial, coronal, and sagittal planes; DVHs; and maximum, minimum, and average target and sensitive structure doses. A description of plan evaluation methods is presented in Chapter 12 "Delivery Systems."

The dose inhomogeneity of an IMRT plan is usually higher than that in 3DCRT as a consequence of increased conformity. Any deviation from a conventional uniform dose scheme should be carefully evaluated to ensure its clinical acceptability. If hot or cold spots are unavoidable, efforts should be made to ensure that they are not located in undesirable locations. For example, a cold spot in the center of the target or a hot spot outside the target should be avoided. Even a hot spot inside the target volume may not be desirable. For example, for prostate cancer, a hot spot close to the urethra is usually not acceptable, particularly if the total dose is escalated. The dose gradient of an IMRT plan near the boundary of the target or OAR can be very high. If the structure(s) is susceptible to the setup uncertainty and/or organ motion, the actual dose received by the target or OAR may be significantly lower or higher than that shown in the plan.^{61,62} In this case, an adequate margin for the structure is important to ensure that the planned dose distribution can be achieved in a clinical setting.⁶³

Beam placement in IMRT is worth discussing. Generally speaking, the beam configuration may have significant influence on the quality of an IMRT plan even when a large number of incident beams (eg, nine beams) are used.⁶⁴⁻⁶⁹ Clinically, however, beam orientations are selected on a trial-and-error basis. To obtain an optimal beam configuration, in principle, one can simply add the degree of freedom of beam angles into the objective function and optimize them together with the beamlet weights.^{65,70} Although this does not pose any conceptual challenge, the computational time becomes excessive because of the greatly enlarged search space and the coupling between the beam profiles and the

beam configurations. The beam intensity profiles have to be optimized for every trial beam configuration because the influence of a set of gantry angles on the dose distribution is not known until beam intensity optimization is performed. A computationally efficient optimization algorithm is necessary to have a clinically practical beam orientation optimization tool. Some progress has been made toward this goal.^{71,72} But before commercial companies implement clinically practical tools for automated or semiautomated beam placement, alternative techniques or even some general guidelines would be useful to facilitate IMRT planning.

One of the appealing approaches is the class-solution method.⁷³ The basic idea is to construct a representative beam configuration based on previous experience for a given disease site and then use this "class-solution" for subsequent treatment planning. Schreiber and Xing systematically investigated the issue and proposed a set of class-solutions for IMRT prostate treatment.⁷⁴ To derive a population-based beam orientation class-solution, a beam orientation optimization algorithm was used to derive the optimal solutions for each individual in a group of 15 patients with prostate cancer. Figure 2-11 shows the distributions of optimal beam angles for five, six, seven, and eight beams for the 15 patients studied. The colored short lines represent directions found in individual cases, and the red bold long lines represent the directions identified as the class-solutions. These results indicate that the beam orientations for a certain incident direction are confined in a certain range and that beam orientation class-solutions may be a reasonable compromise between what is practical and what is optimal for prostate IMRT. For other dis-

ease sites, beam orientation class-solutions may not exist because the geometric variations among the patient population are too large.

On approval of the plan by the physician, an RT plan file or a DICOM-RT file is generated containing all of the relevant machine parameters for IMRT treatment. The IMRT plan file can be complex. For example, it may contain hundreds of MLC segments. Consequently, manual delivery is not an option. Instead, delivery is usually accomplished by the computer-controlled systems, including the record and verify system, linear accelerator control, and MLC control software. The detailed treatment settings contained in the RT plan or DICOM file are transferred from the planning system to the record and verify system. Normally, redundant checksums are also in place for each record, ensuring the safe transfer of data over the computer network. Although it is perhaps not necessary to list all of the information about the treatment in the chart, the chart should contain concise information about the treatment that can be easily verified, for example, the treatment machine, energy, number of beams, gantry and couch angles, monitor units (MUs) for each beam, number of fractions, and fraction doses. The plan output, such as isodose lines for a selected plan in different views on CT images, DVHs, and the QA report, should also be documented. The intensity maps for each beam should be included if possible.

Advanced Topics in IMRT Treatment Planning

Inverse planning is at the foundation of IMRT, and its performance critically determines the success of an IMRT treatment. Unfortunately, the currently available inverse planning formalism is not satisfactory, and the solutions out of so-called "optimization" systems are often suboptimal. Considerable effort may be required to compute a clinically acceptable plan, and the final results may strongly depend on the planner's experience and understanding of the planning system. These shortcomings of the existing systems are familiar to anyone engaged in clinical IMRT treatment planning. In addition to the prescription doses, the current planning system requires the user to preselect the angular variables (gantry, couch, and collimator angles) and the weighting factors of the involved structures. These variables and parameters constitute an additional multi-dimensional space, which is coupled to the beam profiles.

A survey carried out by us indicates that there are five major problems with current inverse planning systems: (1) no effective mechanism for incorporating prior experience into plan optimization; (2) lack of direct control over the regional dose or, more generally, lack of interactive tools to guide the planning process; (3) no effective tools for aiding beam placement in IMRT planning; (4) inability to incorporate organ motion directly; and (5) inefficient inter-

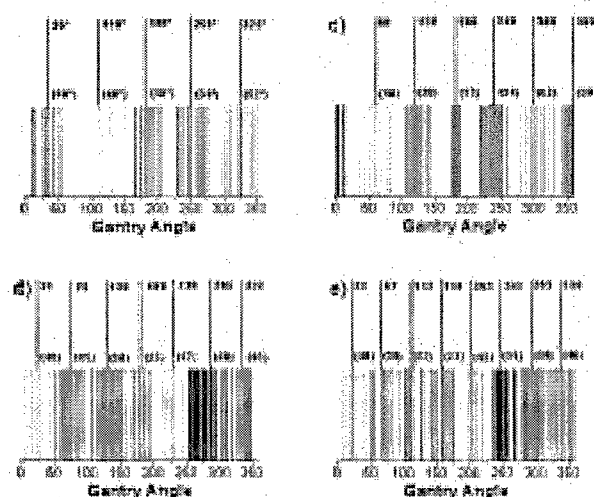


FIGURE 2-11. Distributions of beam angles for five, six, seven, and eight beams, respectively. The short colored lines represent directions found in individual cases, and the long red (bold) lines represent the directions identified as class-solutions. Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

face between planning and delivery systems. Toward establishment of a clinically efficient and robust inverse planning system, many investigators have attacked the problems mentioned above, some of which are the subject of the following sections.

Statistical Analysis-Based Formalism for Therapeutic Plan Optimization

An important element that is missing in the current inverse planning formalism is a mechanism for incorporating prior knowledge into the dose optimization process. In image analysis and many other fields, it has proven valuable to include partial knowledge of the system variables into the optimization process^{35,54,75} because it provides guidance in the search for the truly optimal solution. Statistical analysis formalism, which appears in virtually all branches of the sciences and engineering, affords a natural basis for this type of application and provides a powerful vehicle to achieve the goal of treatment plan optimization. Using this approach, Lian and colleagues demonstrated the feasibility of incorporating a range of prioritized dose prescriptions into the planning process.^{52,76} The approach is based on a newly introduced concept of a preference function, whose role is to relax our requirement of a rigid dose prescription, to allow a range of doses to be considered, and to quantify the willingness to accept a dose in that range. In addition, to make the system less ill-defined, this new scheme can be used to formalize our clinical knowledge (such as outcome data^{19,77}) and incorporate them into dose optimization. In Figure 2-12, we show the preference function derived using published data from Eisbruch and colleagues⁷⁸ for parotid glands (four different irradiation volumes). Coupled with the statistical inference techniques,^{53,54} this should make the inverse planning process more computationally intelligent.

Another application of the formalism is to include model parameter uncertainties into dose optimization. For example, the radiobiologic formalism involves the use of model parameters that are of considerable uncertainty. Biologic "margins" have been used to account for the variability in radiation sensitivity. This method assumes the patient to be more sensitive than the mean value for normal tissues and more resistant than the tumor. EUD-based optimization with the incorporation of model parameters has been demonstrated through the use of a statistical inference technique.⁷⁹ Because currently available models for computing the biologic effects of radiation are simplistic and the clinical data used to derive the models are sparse and of questionable quality, the technique is valuable to minimize the influence of statistical uncertainties.

Multiobjective Optimization

Radiation dose optimization is intrinsically a multiobjective problem because of the existence of multiple conflicting objectives in the system. In the conventional

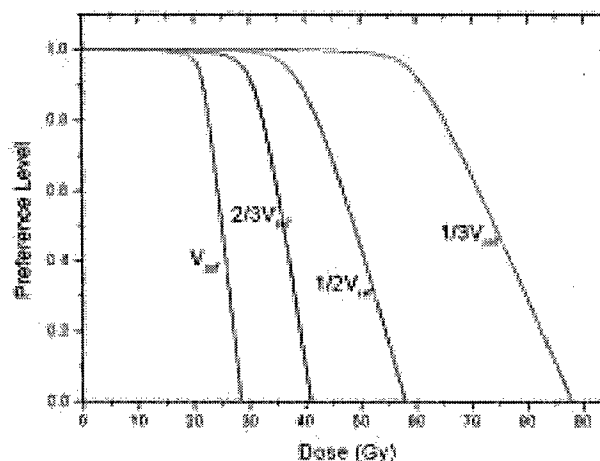


FIGURE 2-12. Preference function of parotid salivary glands for four fixed partial volumes (V_{eff}). In constructing the assumed dose-volume preference function, the preference level is assumed to be 1 for normal tissue complication probability (NTCP) < 5% and 0 for NTCP > 50%. The reference volume (V_{ref}) refers to the volume of the gland. Adapted from Eisbruch A et al.⁷⁸ (To view a color version of this image, please refer to the CD-ROM.)

approach, the multiple objectives are combined to form an overall objective function through the use of so-called importance factors.^{25,80} Contrary to this, the dose delivered to each structure constitutes one of the objectives in multiobjective optimization, which is an alternative way to deal with the trade-offs of multiple conflicting objectives. The method attempts to obtain all efficient solutions and provide the planner with a more thorough picture of the possible options or the trade-offs between the different objectives. Here an efficient solution (often called the Pareto solution in multiobjective optimization theory) is defined as a plan with a good compromise of all of the objectives involved in the problem or, more precisely, a plan that cannot be further improved without significantly deteriorating the dose distribution in one or more organs. Mathematically, the multiobjective optimization (or vector optimization)⁸¹⁻⁸⁴ is to determine a set of decision variables that optimizes a vector function whose elements represent M objective functions without violating the system's constraints. The collection of all efficient solutions is named the Pareto front. Although the approach is conceptually appealing, practical issues, such as the enormous computing time required to obtain the Pareto solution and how to effectively select a plan from the Pareto front, must be resolved before it finds widespread application in RT plan optimization. Perhaps a hybrid of single- and multi-objective techniques is a viable option.

Integration of IMRT Planning and Delivery

IMRT planning is generally performed in two steps: calculation of the intensity maps of the incident beams and decom-

position of each intensity map into a series of MLC-shaped segments using a leaf sequencing algorithm. In practice, the decoupling of dose optimization and leaf sequencing has a number of consequences. In addition to the need for the development of complicated MLC leaf sequencing software, the number of segments resulting from the approach is often unnecessarily large. The leaf sequencing algorithm sometimes has to go through additional steps to accommodate some special hardware constraints of the MLC delivery system that can be easily dealt with at the stage of dose optimization. To improve the efficiency of the interface between the inverse planning and the dynamic MLC delivery systems, attempts have been made to incorporate machine constraints and other physical aspects of the delivery system into dose optimization. The most effective method is perhaps the aperture- or segment-based optimization, which optimizes directly the objective function with respect to the shapes and weights of the segmented fields.^{44,51,85,86} In this approach, the number of segments for each incident beam is prespecified instead of left "floating." Generally speaking, it is more computationally involved to optimize an objective function with respect to the segment shapes and weights because of the nonlinear dependence of the dose on the leaf coordinates. However, the benefits gained by eliminating the extra leaf sequencing step and the associated drawbacks outweigh the slight computational cost.

Interactive Planning Tools for IMRT

The interactive process of IMRT planning is less intuitive than that of forward planning because of the involvement of a large number of parameters whose roles in the final solution are not explicitly known until the completion of a dose optimization calculation. There is a need for the planner to adaptively modify or fine-tune a solution toward the desired direction. For example, frequently after optimization, the dose in only a few small regions is not satisfactory. Currently, plan modification is achieved by adjusting structure-dependent system parameters (eg, prescription, importance factors), which influence not only the dose in the region of interest but also in other areas. To modify the dose in a specific region, in principle, one can use ray-tracing to find the beamlets intercepting the area and adjust their intensities accordingly. The problem is that there are numerous ways to modify this intensity and the optimal arrangement of the beamlet intensities is not obvious. Cortrutz and Xing pointed out that local dosimetric behavior can be more effectively controlled by introducing a region-dependent penalty scheme and demonstrated the utility of this approach using a model system and clinical examples.^{87,88} After the conventional planning is done, they identify the subvolumes on isodose layouts or the dose interval on the DVH curve in which the fractional volume needs to be changed. The local penalty (eg, local importance factor or local prescription) is then adjusted, and the dose is

reoptimized. The fine-tuning of doses is manually iterative in nature, and the process can be easily accomplished using a graphic user interface. Using this technique, it has been shown that one can eliminate hot and cold spots. Generally, in dose optimization, there is no net gain. That is, the improvement in the dose to a region is often accompanied by a dosimetrically adverse effect(s) at another point(s) in the same or different structures. Practically, however, some dose distributions are more acceptable than others. The important issue here is to find the solution that improves the dose(s) at the region of interest with a clinically insignificant or acceptable sacrifice.

It is useful to mention that some "hot spot editor" tools have recently been implemented in commercial systems. These editors rely primarily on a rudimentary ray-tracing and is done as follows: (1) visually locating the hot/cold spot; (2) finding the corresponding beamlets that contribute to the dose at the point of interest (POI); (3) decreasing/increasing the intensities of one or more of the beamlets; and (4) updating the dose distribution. In reality, there are multiple beamlets contributing to the dose at the POI and the problem is determining the optimal way to modify them so that the doses at other points are compromised minimally. In the current commercial systems the hot/cold spot is improved by decreasing/increasing the beamlets with a pre-designed updating method, which is rarely optimal and often causes new cold/hot spots somewhere else within the patient. The approach described by Cortrutz and Xing allows optimal adjustment of the beamlet intensities, thus avoiding the aforementioned problem. Since the re-optimization is done on top of the existing solution, it requires only very limited additional computing time. Furthermore, it is done in the background (just like the recalculation of dose in step (4) is done in the background) and the user does not need to take any additional action.

Automated and Semiautomated Beam Placement

Clinically, gantry angles are selected empirically, and there is no guarantee that the beam configuration is optimal for a given patient. Many investigations are exploring the role of beam configuration selection in IMRT^{39,65,69,89-91} and developing tools for beam placement. A promising technique uses beam's eye view dosimetrics (BEVD).^{71,92} The central idea of this single-beam scoring technique is that the merit of a beam direction should be measured by what that beam could achieve dosimetrically without exceeding the dosimetric or dose-volume constraints of the system. For computational purposes, a beam portal is divided into a grid of beamlets. Each beamlet crossing the target is assigned the maximum intensity that could be used without exceeding the dose tolerances of the sensitive structures and normal tissue. A forward dose calculation using the "maximum" beam intensity profile is then performed, and the score of the given beam direction (indexed by i) is calculated according to⁹²

$$S_i = \frac{1}{N_T} \sum_{\text{all Target}} \left(\frac{d_{ni}}{D_T^p} \right)^2 \quad (6)$$

where d_{ni} is the maximum dose delivered to voxel by the beam from the direction indexed by i , N_T is the number of voxels in the target, and D_T^p is the target prescription. The BEVD score function captures the main feature of a planner's judgment about the quality of a radiation beam and allows one to select beam orientations without excessive computational time. For a given patient, the score function for every possible beam direction is evaluated and the directions with the highest BEVD scores are identified. Although the technique does not yield the final beam configuration in a multifield IMRT treatment, it provides useful information with regard to which beam directions are potentially good or bad. During planning, the beams with the highest scores are considered favorable for the treatment. It is also illustrative to point out that the BEVD information can also be integrated into beam orientation optimization program to improve the convergence behavior and computational speed.⁷²

Hybrid Treatment of IMRT with Other Modalities

IMRT affords one the ability to produce not only spatially uniform but also purposely nonuniform doses. A natural application of the feature is to combine IMRT with other RT modalities to generate a dose distribution that would otherwise be impossible. Along this line, IMRT has been considered a method of salvaging suboptimal prostate implants.⁹³ The combination of IMRT with conventional electron beam(s) for improving the photon-electron field matching and for treatments of certain specially shaped targets has also been investigated.⁹⁴ Figure 2-13 illustrates a combined head and neck boost treatment using electron and intensity-modulated photon beams. In head and neck cancer, treatment initially involves the irradiation of the primary tumor and the cervical lymph nodes. After the tolerance dose of the spinal cord is reached using opposed lateral photon beams, the lateral fields are reduced off the spinal cord. The treatment of the anterior neck along with the primary tumor is continued using the reduced photon beams, whereas the posterior neck overlying the spinal cord is treated with lateral electron fields. This requires the matching of an electron field with two opposed photon fields. For comparison, the combined treatment using the conventional technique (electron + conventional photon beams) is shown in Figure 2-13. When the electron and unmodulated photon beams are matched directly, hot spots greater than 15% above the prescribed dose are seen in the abutting region. This is reduced to 5% when the proposed technique was used, with markedly better dose homogeneity in the abutting region. Furthermore, because of the broadened photon penumbra, the dose homogeneity in the junc-

tion region becomes less sensitive to patient setup errors.⁹⁴

Gated or Synchronized IMRT

IMRT can produce highly conformal doses to targets and a sharp dose gradient between targets and surrounding critical structures.^{60,95-98} Together with improved patient immobilization, the target margins can be reduced to facilitate dose escalation. However, margin reduction remains challenging for treatments in the thoracic regions and other sites in which intrafraction respiratory motion is significant.^{99,100} The delivery of IMRT in either dynamic or static mode can cause unexpected high- and low-dose regions owing to the interplay between the movements of the tumor and the MLC leaves.¹⁰¹ This issue can, in principle, be improved through

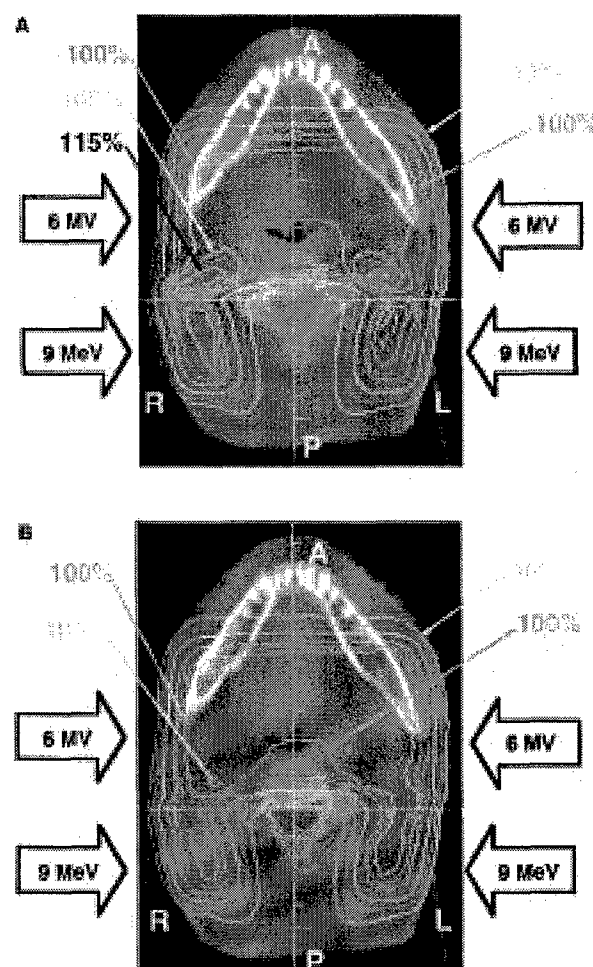


FIGURE 2-13. Comparison of the isodose distributions of the treatment plans in a transverse section of a patient with head and neck cancer when the electron and photon beams were matched directly (A) and when dynamic intensity modulation was used for the photon beams (B). Isodose levels are shown at 30%, 50%, 70%, 90%, 95%, 100%, 105%, 110%, and 115%. Note that the 110% and 115% isodose lines are not present in (B). Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

the use of gating or respiration synchronization.¹⁰²⁻¹⁰⁶ (See Chapter 9 "Respiratory Motion Management" and Chapter 19.4 "Intrafractional Organ Motion and Planning: Emerging Technology".)

In gating, tracking, and breath-hold, the treatment machine is switched "on" or "off" in response to a signal that is representative of a patient's breathing motion. Both passive and active devices can be used to monitor the respiratory motion. Ideally, the beam is on only during portions of the breathing cycle when motion is small. The disadvantage of this technique is the prolonged treatment time compared with that of nongating approaches.

Motion-synchronized RT is based on two assumptions: (1) tumor motion is considered to be predictable, and the model of motion can be established prior to the treatment and is assumed to be the same (or at least adaptively predictable) throughout the treatment course, and (2) the treatment delivery system, either the MLC leaves or the treatment couch, can be instructed to precisely move to certain locations to adapt to the motion of the tumor. The main advantage of the motion-synchronized RT is that the radiation beam is on all of the time; therefore, there is no treatment time prolongation. However, several major technical difficulties must be overcome. Inverse treatment planning in this case must take into account the functionality of the delivery system.

Another issue is that respiratory motion exists in all stages of the RT process, including preplanning imaging and treatment planning and delivery. If respiratory motion is not accounted for during image acquisition, artifacts may arise during the image acquisition, leading to the distortion of the target volume. For gated treatment, the same window should be used for imaging and planning and delivery so that tumor positions and patient anatomy can be reproduced accurately. For motion-synchronized RT, several sets of CT images representing different phases of the breathing cycle need to be acquired through either a high-speed multislice CT scanner or by postprocessing software to sort the images. All of these images will be used for IMRT treatment planning, and the resulting MLC leaf sequences need to be multiplexed for delivery.

Biologically Conformal IMRT

Although the biology of tumors plays a crucial role in the success of RT, commonly used CT and magnetic resonance images provide few metabolic data and have significant shortcomings in characterizing benign and malignant tumors. Recent advancements in functional imaging make it possible to noninvasively obtain a patient's metabolic distribution. Coupled with the technical capability of IMRT in generating customized three-dimensional dose distributions with subcentimeter resolution, this may afford a significant opportunity to improve conventional RT by producing doses in accordance with biologic requirements.¹⁰⁷⁻¹¹⁰ Research effort is focused on integrating functional data into IMRT treatment planning to improve clinical cancer management.

In general, functional imaging suggests nonuniform dose distributions to meet the heterogeneous biologic requirements. Xing and colleagues identified some relevant issues and developed a preliminary four-dimensional inverse planning scheme for functional imaging-guided IMRT.¹⁰⁹ The metabolic and functional data are incorporated effectively by modulating the prescription doses in the target voxels. This algorithm enables one to produce a high dose where there is resistance and/or where tumor burden is large and to differentially spare the sensitive structures with more emphasis on functionally important regions.

Delivery Techniques for Fixed-Gantry IMRT

IMRT delivery with MLC is based on the simple principle that moving jaws or leaves can be employed to control the dose delivered to a point. As mentioned previously, IMRT planning is currently performed in two steps: optimization of intensity maps and MLC leaf sequencing. The latter is to convert an intensity map into an MLC leaf sequence file, which specifies the leaf positions as a function of the fraction of MUs delivered. For the delivery, the two-dimensional beam fluence is divided into strips corresponding to the projection of each leaf pair of the MLC. Each MLC leaf pair is then required to modulate the fluence along its projection (see Chapter 12 "Delivery Systems"). For Varian linear accelerators (Varian Medical Systems, Palo Alto, CA), the leaf pairs are independent, reducing the conversion of two-dimensional fluence profiles into a collection of one-dimensional problems. As a result, the problem becomes finding a series of leaf positions (coordinates of leading and trailing leaves) to cover the area under a one-dimensional fluence function. There is no unique solution to this problem, leading to a number of ways to accomplish beam modulation.

MLC-based delivery is generally divided into static step-and-shoot¹¹¹⁻¹¹⁶ and dynamic modes.¹¹⁷⁻¹¹⁹ A step-and-shoot leaf sequence file consists of alternatives of dose-only and motion-only instances. The step size of MLC movement in this mode is determined by the dimension of the beamlet in the leaf movement direction. Dynamic delivery differs from the step-and-shoot mode in that leaf movement and dose delivery are realized simultaneously. These algorithms are described below.

Step-and-Shoot Delivery

In step-and-shoot delivery, the total dose at a spatial point is the superposition of contributions from a series of segment fields (typically, the number of segments is between 20 and 100). The x-ray beam is off when the MLCs travel from one segment to another. This is perhaps the most intuitive technique to deliver intensity-modulated fields using MLC. The QA procedure for this delivery mode is relatively simpler (than

dynamic delivery) because there is no correlation between the leaf speed and the dose. Instead of describing the algorithm generally, an example is used to illustrate how the step-and-shoot leaf sequencing methods work.¹¹¹

Figure 2-14 depicts a simple example of an intensity pattern. The intensity in a 6×4 cm field is expressed using five discrete intensity levels. Four 1 cm-wide leaves (numbered 12, 13, 14, and 15) are to be used to generate the intensity pattern. The profiles that each leaf pair must generate are shown in the individual graphs (the profile required by leaf pair 12 in this example).

For a modulated field to be delivered at a gantry angle, each component profile along the center of the j^{th} leaf pair must be rendered into a leaf trajectory. Intensity modulation along the profile, as shown in Figure 2-15, is obtained by sweeping the leading leaf, 12B, and the following leaf, 12A, from left to right along the x-axis. The first step in this procedure is to divide the total relative beam intensity into a number of equal intervals of width $\Delta\Phi$, as indicated in the illustration. The number of intervals selected to span the range of the intensity is NI. The second step in the procedure is to find the intersection of the centers of these profile increment bins with the profile. These points are

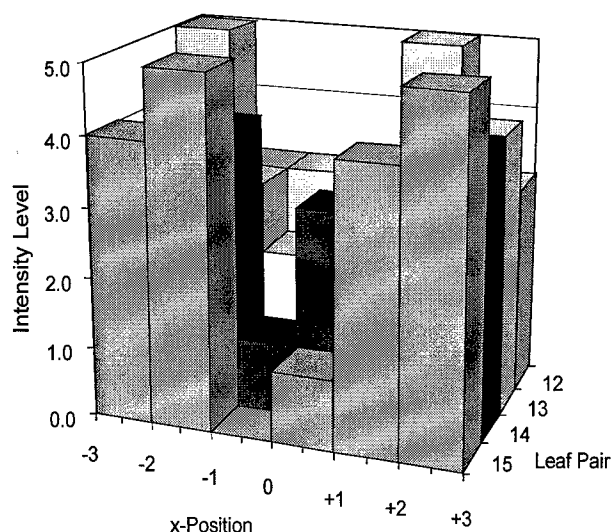


FIGURE 2-14. Intensity map used for illustrating the step-and-shoot leaf sequencing algorithm. (To view a color version of this image, please refer to the CD-ROM.)

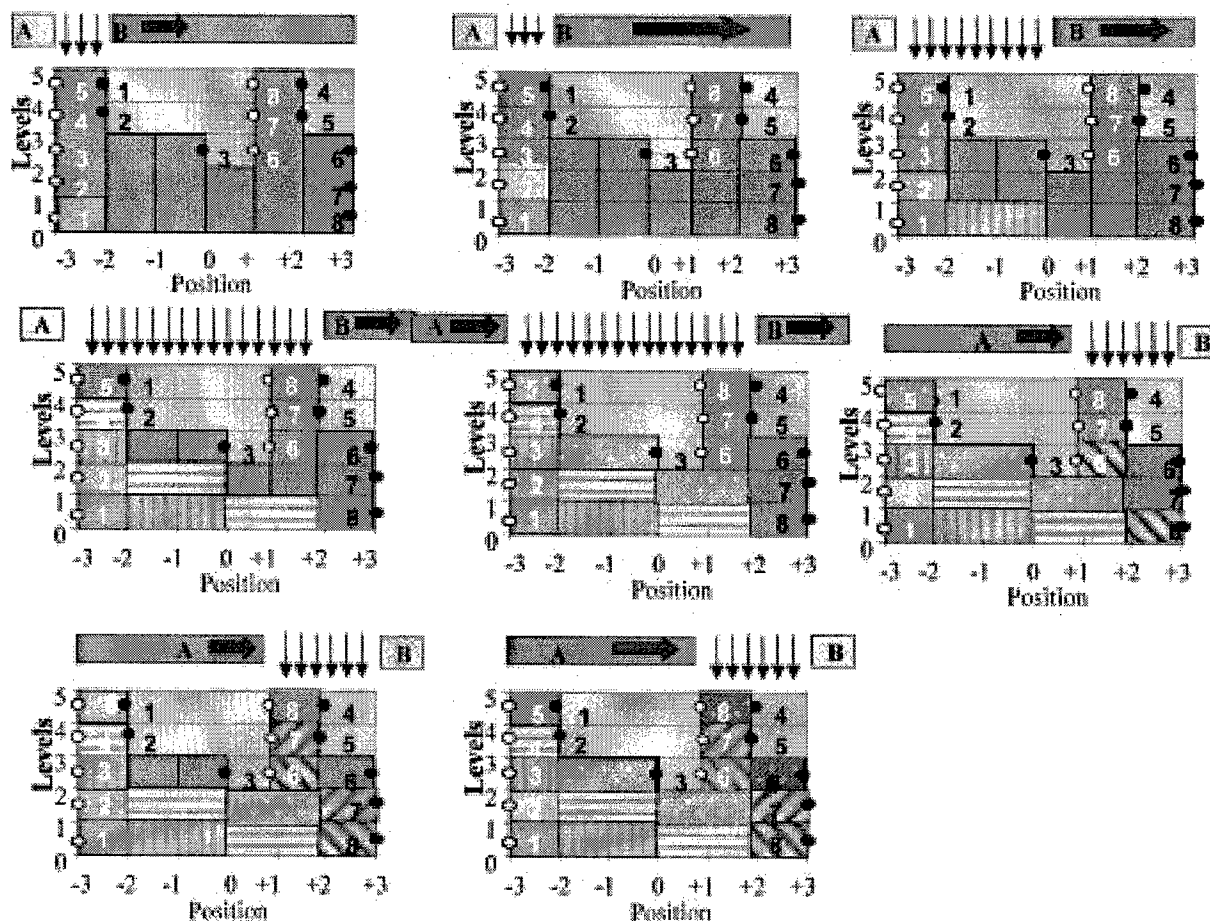


FIGURE 2-15. Intensity profile to be produced by leaf pair 12. Reproduced with permission.

indicated by circles in Figure 2-15. The algorithm requires that an even number of such points be found. The third step is to divide the coordinate points into two groups. One group consists of those points lying on an ascending slope of the profile in which there is a positive gradient (open circles in Figure 2-15), and the other group consists of those points lying on a descending slope of the profile in which there is a negative gradient (filled circles). The fourth step is to rank the points in each group. The numbers indicated are the i -index for the sequence for the twelfth pair of leaves. Pairing together the coordinates of equal rank order and assigning the coordinates to each pair of leaves produces the desired leaf sequence for the k^{th} gantry angle position, $\{xA_{i,j,k}, xB_{i,j,k}\}$, where the index i ranges from 1 to NI_j . The number of steps required to create the trajectories will not be the same for all profiles that make up a field. Steps must be added to the shorter sequences with the leaves abutting beneath a jaw at one end of the profile so that all sequences for a field will have the same number of steps.

Another type of step-and-shoot delivery is based on the sequential reduction of intensities according to a prespecified scheme.^{112,113,120} The pattern of integers in Figure 2-16 represents an intensity pattern to be delivered using this leaf

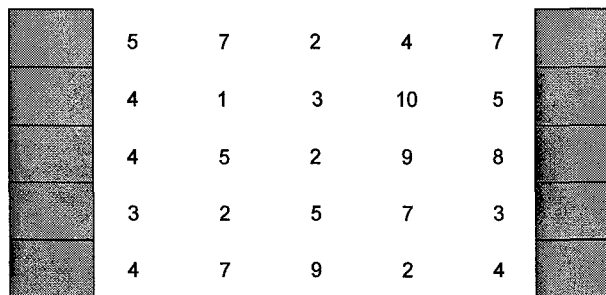


FIGURE 2-16. Example intensity map used for illustrating the "areal" leaf sequencing algorithm.

setting sequence. The 5×5 cm field is to be delivered with a maximum beamlet intensity of 10 and a minimum beamlet intensity within the field of 1. The underlying principle of the algorithm for determining the sequence is that the most efficient way to subdivide a sequence is by halves. The sequence is to be delivered by increments that are powers of 2. In this case, the increments are 8, 4, 2, and 1. The first step is to set the leaves in a pattern that can deliver an exposure of 8. There are four beamlets with intensities of 8 or more. They are not contiguous, but leaves can be set to form two windows around the two regions that each deliver an intensity of 8. This is step 1 in Figure 2-17. After this exposure, all but one of the beamlet positions still require an exposure of 1 or more. A leaf pattern can then be found that exposes beamlets that require a residual exposure of 4 or more. However, two such regions exist that require two separate sets of leaf settings. These are steps 2 and 3 in Figure 2-17. The residual intensity then contains values up to 3, which can be reduced by exposures of 2. Again, to expose all of the beamlets, two leaf patterns are required, each delivering exposures of 2. These are steps 4 and 5 in Figure 2-17. Then all of the beamlet positions have either received their full exposure or have a residual value of 1. Two more leaf patterns are required to reach all of the 1 positions and reduce the residual intensity to zero. In all, seven steps are required to deliver the intensity pattern. The single-profile step-and-shoot leaf setting algorithm requires 13 steps to deliver this pattern.

Dynamic Delivery

Let $\Phi(x)$ be the fluence along the trajectory of the leaf pair. An example profile is shown in Figure 2-18. To deliver the fluence, one must determine the arrival times at x , $t_A(x)$ for leaf A and $t_B(x)$ for leaf B. The units of the arrival times can be seconds, or they can be expressed as MUs. The irradiation time interval at x between the opening of the ray by leaf B and the shielding of the ray by leaf A is indicated by

Leaf Setting Sequence

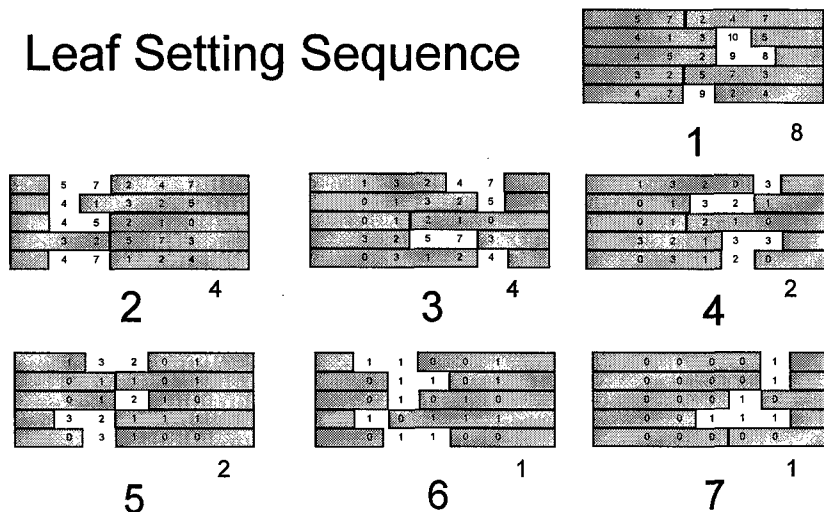


FIGURE 2-17. Leaf sequencing steps involved in decomposing the intensity map shown from Figure 2-16. The "areal" algorithm described by Xia and Verhey¹¹² is used here. Reproduced with permission.

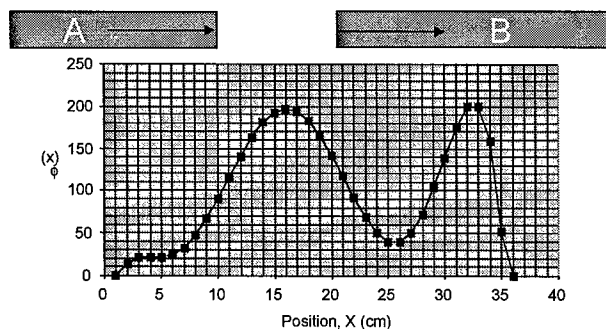


FIGURE 2-18. Example intensity map used for illustrating the dynamic leaf sequencing algorithm. Reproduced with permission.

$$\tau(x) = t_A(x) - t_B(x) = \Phi(x)/\Phi_0. \quad (7)$$

Figure 2-19 can then be considered to be a time-position graph for the two leaves. The upper border of the shaded area is the leaf A trajectory, and the lower border, or x -axis, is the leaf B trajectory. The problem with this interpretation is that it requires leaf B to travel with infinite velocity and leaf A to travel backward in time! The dilemma can be resolved by applying a sequence of operations that transform the two trajectories such that they become deliverable. Note that there are four regions marked along the fluence profile in which the gradient is either positive or negative. To remove the time reversal from the continuous fluence profile, a reflection operator is introduced and is defined by

$$\tau' = \Delta \tau_{R1} \pm [\tau(x) - \Delta \tau_{R1}] \quad (8)$$

$$\tau' = \Delta \tau_{R2} \pm [\tau(x) - \Delta \tau_{R2}] \quad (9)$$

where $\Delta \tau_{R1}$ is the average value of the portion of the profile with a negative gradient within the R1 region around the first maxima and $\Delta \tau_{R2}$ is the average value of the portion of the profile with a negative gradient within the R2 gradient region around the second maxima. The positive sign is applied when there is a positive gradient and the negative sign when there is a negative gradient. The reflection operator is applied to the curves in the negative gradient regions to yield curves that do not require the leaves to travel backward in time. The results are shown in Figure 2-22.

The operations have introduced a discontinuity in the leaf sequence curves that can be removed by applying a translation operator defined in region R1 by

$$\tau''(x) = \tau'(x) + \Delta \tau \quad (10)$$

where the increments are selected to remove the discontinuity between region R1 and R2, as illustrated in Figure 2-20. For the sake of generality, in region R1, the translation constant is zero. However, now there are still horizontal portions of the curves that represent infinite velocity of the leaves. There is always a horizontal segment occurring in either leaf A or leaf B trajectories across the entire sequence.

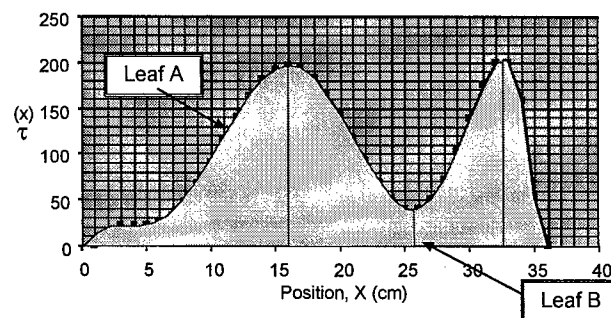


FIGURE 2-19. Time-position graph for the two leaves during the dynamic delivery process. Reproduced with permission.

To remove the infinite velocity, some additional slope is introduced to each leaf trajectory. This can be achieved by applying a shear operator to the entire lengths of both leaf trajectories. The shear operator is defined by

$$\tau'''(x) = \tau''(x) + x/v_{\max} \quad (11)$$

This operator tilts the upper and lower horizontal bounds of each segment of the sequence by an amount determined by the maximum leaf velocity, resulting in a sequence that can be practically delivered. The slope of the shear is the inverse of the maximum velocity that the leaves can move, v_{\max} . The resulting leaf setting sequence is shown in Figure 2-20. The leaves begin the sequence closed at the left side of the field and end the sequence closed together at the right side of the field. In a region in which the original fluence gradient is positive, the leading leaf, leaf B, moves with a constant speed determined by the maximum velocity, and the trailing leaf A moves along the trajectory given in equation 11. In those regions in which the fluence gradient is negative, the trailing leaf, leaf A, moves with the maximum velocity, whereas the leading leaf moves along the trajectory given in equation 11.

The algorithm used to calculate the velocity modulation of the slower leaf can be derived by differentiating equation 12 with respect to distance:

$$d\tau'''/dx \equiv 1/v(x) = d\tau''/dx + 1/v_{\max} \quad (12)$$

The derivative of τ'' with respect to x can be obtained from equation 10 and is simply the derivative of τ' in all subdivisions of the trajectory. The derivative of τ' can be obtained from equations 8 and 9 and depends on the sign of the fluence gradient:

$$\nabla \Phi < 0 \Rightarrow d\tau'/dx = -d\tau/dx \quad (13a)$$

$$\nabla \Phi > 0 \Rightarrow d\tau'/dx = +d\tau/dx \quad (13b)$$

The derivative of τ can be seen in equation 7 to be

$$d\tau/dx = (d\Phi/dx)/\Phi_0 \quad (14)$$

assuming that the variation in the incident fluence is negligible with respect to x . Using these results in equation 12, the velocity modulation equation becomes

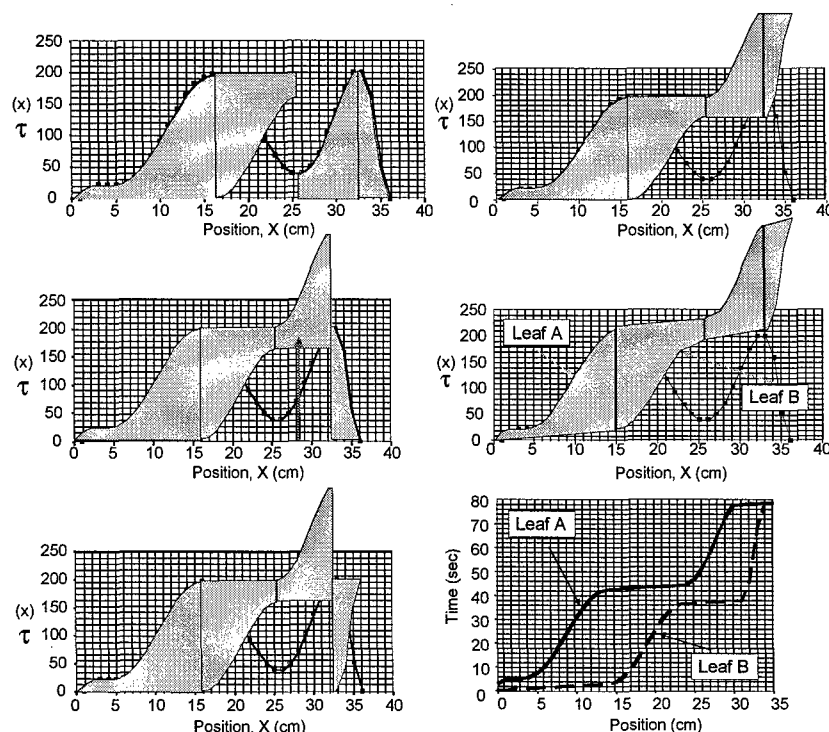


FIGURE 2-20. Trajectories of leaf A and leaf B during the dynamic delivery process. Reproduced with permission.

$$1/v(x) = \pm (d\Phi/dx)/\Phi_0 + 1/v_{\max} \quad (15)$$

where the positive sign applies to positive fluence gradient regions and the negative sign applies to negative fluence gradient regions. By rearranging, one arrives at

$$v(x) = v_{\max} / [1 \pm v_{\max} \cdot (d\Phi/dx)/\Phi_0] \quad (16)$$

This equation can be used to generate the velocity modulation required to deliver the fluence profile, starting with the leaves closed together at one side of the profile and ending with the leaves closed together again at the other. The leaf setting sequence computed by the velocity equation for the original fluence in Figure 2-18 is shown in Figure 2-20. The results are exact.

Our experience with both step-and-shoot and dynamic delivery indicates that there is no clear-cut advantage for any one of the methods except in some special situations. The main disadvantage of the step-and-shoot method is a sacrifice of accuracy in the delivery of beam profiles that have steep gradients. The dynamic method delivers the required distribution by sweeping the leaf pair across the beam and becomes inefficient in producing "large and flat" fluence segments.¹²¹ It is possible to implement an algorithm combining the step-and-shoot and the dynamic deliveries to use the advantages of each.¹²² This scheme would determine the slope of each segment of the intensity profile and then choose the suitable delivery method.

Related Issues in MLC-Based Delivery

Unlike conventional RT with static MLC fields, significant dosimetric issues must be addressed when IMRT delivery is used. Most of the algorithms in the literature, however, assume an ideal MLC and ignore the influence of many physical effects and the mechanical constraints of a realistic MLC, such as transmission and head scatter, tongue-and-groove effects, and collision constraints for adjoining leaf pairs.

Yang and Xing proposed an algorithm to account for the leaf transmission and head scatter effects in step-and-shoot leaf sequencing.¹²³ In their approach, an error function, defined as the least square difference between the desired and the delivered fluence maps, is introduced. Mathematically, this function is expressed as

$$F = \sum_{i,j} [\varphi(i,j) - \varphi_d(i,j)]^2 \quad (17)$$

where $\varphi(i,j)$ and $\varphi_d(i,j)$ are the calculated and the desired fluences of beamlet (i,j) , respectively. In equation 18, only those beamlets with nonzero fluences in the desired intensity map are considered because one cannot physically produce a beamlet with zero fluence.

The calculation starts with the MLC leaf sequence file derived from the desired fluence map without considering

transmission and head scatter. The effects of transmission and head scatter are minimized by iteratively adjusting the fractional MUs in the initial MLC leaf sequences using a downhill simplex optimization method. A three-source model¹²⁴ is used to evaluate the relative head scatter contribution for each segment. The three effective sources are the source for the primary photons from the target and two extrafocal photon sources for the scattered photons from the primary collimator and the flattening filter, respectively. The algorithm has been assessed by comparing the dose distributions delivered by the corrected leaf sequence files and the theoretical predication, calculated by Monte Carlo simulation using the desired fluence maps and several clinical IMRT cases. The deviations between the desired fluence maps and the ones calculated using the corrected leaf sequence files are less than 0.3% of the maximum MU for the test field and less than 1.0% for the clinical IMRT cases. The experimental data demonstrate that both absolute and relative dose distributions delivered by the corrected leaf sequences agree with the desired ones within 2.5% of the maximum dose or 2 mm in high-dose gradient regions. It is found that the influence of the two effects is more pronounced in the absolute dose than in the relative dose. Figure 2-21 illustrates a measured absolute dose profile for a test field. In performing the measurements, MLC leaf sequences with different correction schemes described in the figure caption were used.

The influences of rounded leaf ends and interleaf transmission,¹²⁵ tongue-and-groove effect,¹²⁶ and the effect of back-scattered photons from the moving jaws and MLC leaves on the monitor chamber signal¹²⁷ have been studied using accurate models with realistic MLC geometries. It has been pointed out that the tongue-and-groove effect may be significant when underdosage occurs between two adjacent leaf pairs owing to the fact that the region between is always covered by the tongue, the groove, or both.^{128,129} Algorithms

have been developed to either minimize or remove this effect when MLC leaf sequences are generated.^{114,130-133} Many researchers have shown that the tongue-and-groove effect can result in an underdose of as much as 10 to 15% in some special situations.^{128,129,134,135} However, a Monte Carlo study by Deng and colleagues suggested that the difference between the dose distributions with and without the tongue-and-groove effect was hardly visible for an IMRT treatment with multiple gantry angles in a clinical setting.¹²⁶ More thorough investigations on the tongue-and-groove effect and other physical factors are needed to understand the system and to determine better solutions.

Finally, the inclusion of various physical factors is made simple if segment-based inverse planning is used. This represents one of the major advantages of the new type of inverse planning approach with integration of machine constraints.

Quality Assurance

IMRT adds a new degree of freedom to conventional RT and allows one to tune the dose distribution on an individual beamlet level. At the same time, it significantly increases the level of sophistication and complexity of the planning and delivery systems. With more and more institutions starting IMRT programs, it becomes increasingly important to have robust and efficient QA tools for clinical use. Otherwise, the gain from IMRT may be lost in a nonoptimal QA procedure and/or be offset by the increased cost of treatment. In general, IMRT QA has three aspects: commissioning and testing of the inverse treatment planning and IMRT delivery system, routine QA of the MLC delivery system, and patient-specific validation of each treatment plan. The first task is mainly concerned with the integrity of the IMRT system. The second involves the normal operation of the dynamic delivery system, and the third

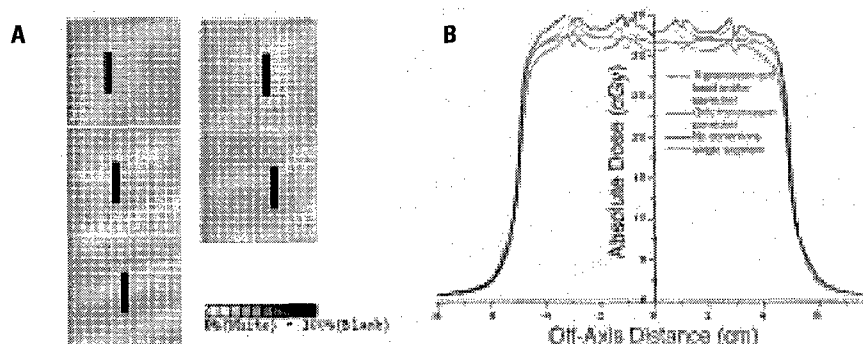


FIGURE 2-21. (A) A schematic diagram of the test field. The field includes five consecutive $2.0 \times 10 \text{ cm}^2$ segments and attempts to produce a $10 \times 10 \text{ cm}^2$ open beam. (B) The measured absolute dose profiles along the midline of leaf pair 21A-21B in the isocenter plane at a depth of 5 cm in solid water for the test field are shown on the left. The measured results of the single-segment $10 \times 10 \text{ cm}^2$ open field with 30 monitor units are shown in the central black curve as a benchmark. The red curve is obtained with correction of head scatter and transmission. The top black and bottom blue curves represent the calculated dose profiles with only head scatter or transmission considered.¹²³ Reproduced with permission. (To view a color version of this image, please refer to the CD-ROM.)

task ensures accurate and safe treatment of the patient. Recently, there have been many excellent reports on IMRT QA-related issues.^{60,136-139} Some practical aspects of IMRT QA are also discussed in Chapter 13 "Commissioning and Dosimetric Quality Assurance" and Chapter 14 "Quality Assurance Processes and Future Directions". In this section, the QA procedure and some recent advancements are summarized.

Commissioning and Testing

To ensure that the system can be used safely and accurately, the inverse treatment planning system must be commissioned prior to clinical use. Commissioning and testing consist of four separate but related steps. The first is concerned with the system's ability to accurately compute series of broad beam data. This type of testing is rudimentary but useful to identify potential problems quickly. The second study tests the dose model and the delivery system with several specially designed intensity patterns. The accuracy of dose calculation for intensity-modulated beams can thus be assessed. The third type of study examines the system's functionality and dosimetric correctness for a number of hypothetical phantom cases. In addition to dose calculation, the functionality of dose optimization is evaluated at this level of tests. Figure 2-22 shows two examples of this type of measurement using a cylindrical water phantom and ion chamber.¹⁴⁰ The last type of study is to test the system using clinical cases to ensure the dosimetric accuracy and integrity of the system. This study evaluates the combination results of image acquisition and segmentation, geometric and dosimetric calibration of the planning system, planning and dose calculation, and data transfer. The dose distributions for single or multiple fields are usually done using an ion chamber and films in a phantom. Other dosimeters, such

as thermoluminescent dosimeters and semiconductor detectors, can also be employed. The American Association of Physicists in Medicine (AAPM) Task Group 40 and Task Group 53 reports provide guidelines on this topic and remain the benchmark documents on the subject.^{141,142} This subject has also been discussed extensively in recent publications.^{136,137,143}

Routine Machine QA

Intensity modulation is achieved with computer-controlled MLC using either static^{112,144-146} or dynamic delivery techniques. To ensure that the planned dose distributions are safely and accurately delivered, an important requisition is the normal operation of the delivery system, which is warranted by routine machine QA. The principles and practice of QA for RT can be found in the classic documents of Van Dyk and Purdy,¹⁴⁷ as well as the report of AAPM Task Group 40.¹⁴¹ For IMRT, several things specific to the IMRT MLC control system need to be checked periodically. Currently, the routine accuracy check of MLC leaf positioning in most clinics is performed using radiographic films with specially designed MLC leaf sequences.^{148,149} Besides being time consuming, the results of film measurements are difficult to quantify and interpret. A few research groups have attempted to use an electronic portal imaging device (EPID) for quantitative verification of MLC leaf positions with edge detection algorithms.¹⁵⁰⁻¹⁵³ The detection precision is limited to ~ 1 mm owing to the finite pixel size and the signal-to-noise ratio of the EPID, which is clearly insufficient for routine QA of the MLC delivery system. Here we describe a quantitative technique¹⁵⁴ for MLC leaf positioning QA developed recently at Stanford University. Given its simplicity, efficiency, and accuracy, we believe that the technique is ideally suitable for routine MLC QA and should have widespread clinical application in the future.

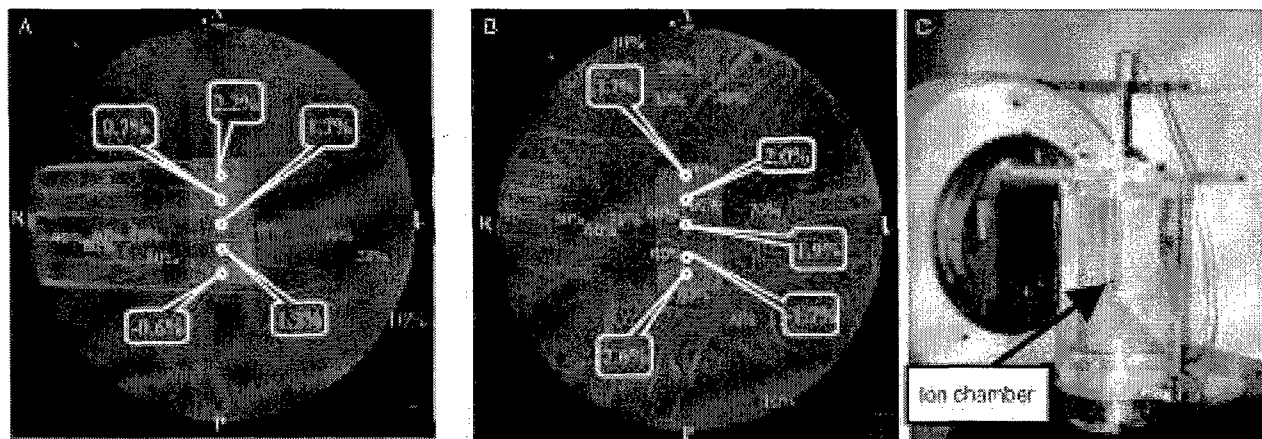


FIGURE 2-22. Hypothetical intensity-modulated radiation therapy plans generated for a cylindric water phantom and the measured dose distributions. The phantom is positioned with its axis perpendicular to the couch top and is supported by a bearing, allowing for rotation about its axis. This allows for measuring the dose of a multifield plan without gantry rotation. The measurements were made using the Varian dynamic multileaf collimator modulating 4 MV x-ray beams. The plans were generated using the CORVUS system. Reproduced with permission from Xing et al.¹⁴⁰ (To view a color version of this image, please refer to the CD-ROM.)

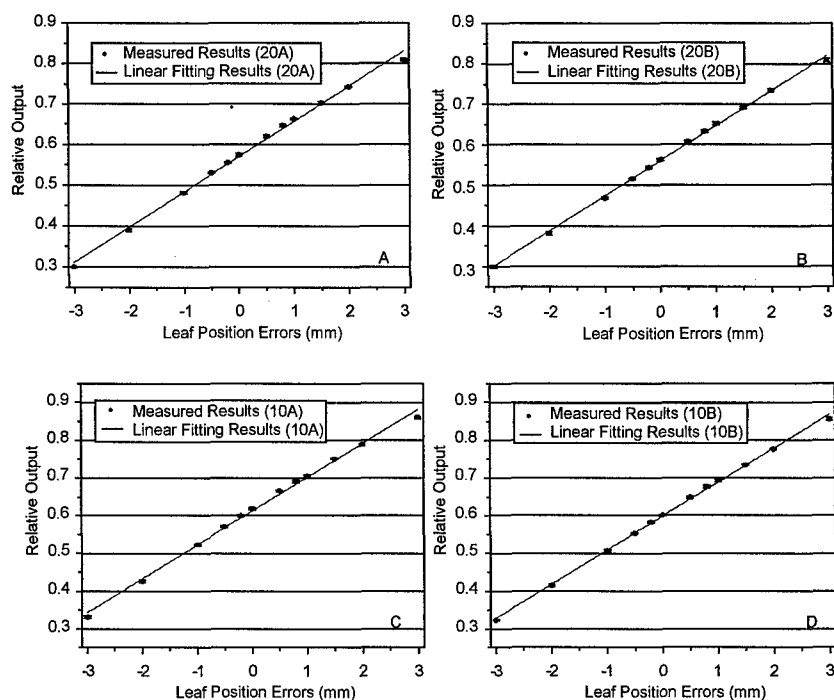


FIGURE 2-23. Relative output versus the displacement of leaves 20A, 20B, 10A, and 10B from their desired positions. The symbols are the measured data, and the solid lines are the least square fitting of the corresponding data sets. Reproduced with permission from Yang Y and Xing L.¹⁵⁴

The Stanford MLC QA technique uses the fact that when a finite-sized detector is placed under a leaf, the relative output of the detector will depend on the relative fractional volume irradiated. A small error in leaf positioning would change the fractional volume irradiated and lead to a deviation of the relative output from the normal reading. For a given MLC and detector system, the relationship between the relative output and the leaf displacement can be easily established through experimental measurements and used subsequently as a quantitative means for detecting possible leaf positional errors. Figure 2-23 illustrates a set of calibration curves for different leaves obtained using an ion chamber and a Varian CL 2300C/D accelerator with an 80-leaf MLC.¹⁵⁴ Our results indicate that the method could accurately detect a leaf positional change of ~ 0.1 mm. The principle of the method is independent of the type of MLC and detector. The method overcomes the previously stated shortcomings of both film measurement and edge detection techniques and provides a reliable means for quantitative examination of MLC positional accuracy.

The principle has also been applied to MLC leaf positioning QA using an EPID,¹⁵⁵ which has the advantage of simultaneously detecting positional errors of any leaf at any point. In this technique, the active imaging region of an EPID is divided into a number of small rectangular regions of interest, each of which is centered at a point at which the leaf positioning accuracy is to be examined (Figure 2-24). Every region of interest here acts as a finite-sized detector,

and the integral signal from it can be processed based on the pre-established relation between the integral signal and the leaf displacement at the point. The EPID-based system also allows us to take the dosimetric influence of the adjacent leaves into account. For this purpose, the integral signal at a region of interest is expressed as a weighted sum of the contributions from the displacements of the leaf above the point and the adjacent leaves. The linear coefficients of the system equations are determined by fitting the integral signal data for a group of predesigned MLC leaf sequences to the known leaf displacements that are intentionally introduced during the creation of the leaf sequences. Once the calibration is done, the system can be used for routine MLC

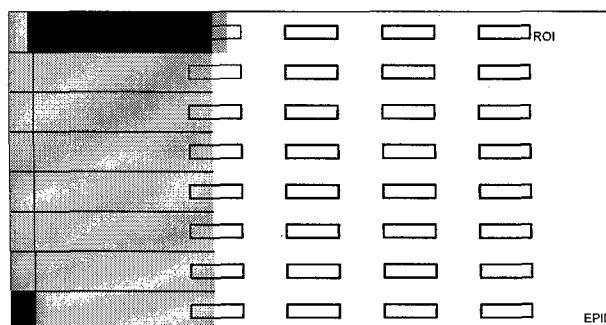


FIGURE 2-24. Diagram of the geometric setup using an electronic portal imaging device (EPID) to examine leaf positioning accuracy. (To view a color version of this image, please refer to the CD-ROM.)

leaf positioning QA to detect possible leaf errors. Table 2-1 shows a set of test data obtained using the technique. Overall, our results show that the proposed technique is superior to the conventional edge-detecting approach in two aspects. First, it deals with the problem in a systematic approach and allows one to take into account the influence of the adjacent MLC leaves effectively. Second, it has a much higher signal-to-noise ratio and is thus capable of quantitatively measuring extremely small leaf positional displacements. The technique can effectively detect a relative lead positional error as small as 0.1 mm at an arbitrary point within the field in the absence of an EPID setup error and 0.3 mm when this uncertainty is considered.

IMRT Treatment Plan Validation

The tasks of patient-specific QA can be divided into geometric and dosimetric verification. The former is concerned with the geometric accuracy of the IMRT beams, including isocenter and portal verification. The dosimetric verification includes a quantitative check of fluence maps, radiation doses at multiple points, and, in some cases, the dose distribution. Currently, the dosimetric verification is primarily done experimentally.

Geometric Verification

A pair of orthogonal simulation films (or digital reconstruction radiographs [DRRs]) is used to verify the patient position by comparison with portal films. In 3DCRT, a portal image is taken using the double-exposure technique, one with the customized radiation port and the other with

a larger rectangular open field, so that both the field boundary and selected patient anatomy can be visualized. A simulation image for an IMRT field can be created as well using the MLC boundary as the port of the radiation field. An example of such a portal image for an IMRT head and neck treatment is shown in Figure 2-25. For portal image exposure, an MLC field that defines the field boundary needs to be extracted from the IMRT leaf sequence file. The MLC-defined field aperture can be appended to the DRR to be displayed together with the patient's anatomy. The DRR in the beam's eye view, as shown on the left in Figure 2-25, is used as a reference for comparing with the portal image for target localization during the treatment.¹⁵⁶

Dosimetric Verification and Independent Dose and Fluence Map Calculations

No consensus has emerged regarding what dosimetric quantities need to be examined to validate an IMRT treatment plan. Patient-specific dosimetric QA typically consists of dose measurements at multiple points and fluence map measurements. Some institutions also perform film dosimetry for each patient treatment. Because of the inherent complexity of the problem, it may be some time before definitive recommendations come from national organizations. In general, the goal of the dosimetric verifications is to ensure that the delivered dose distribution agrees with the one from the treatment planning system. The descriptions on equipment and procedure for these measurements have been the subject of a few recent review articles. The fundamental philosophy of IMRT QA and our experience with computer-based patient-specific QA are presented here.

First, one should note that the 3DCRT approach based on point dose verification is insufficient to validate an IMRT plan because of the independence of the involved beamlets. In 3DCRT, verification is mainly concerned with the MU calculation for each incident field. An independent

TABLE 2-1. Detected Leaf Positional Errors with Different Intentionally Introduced Errors for Different B-Bank Leaves at Two Locations

Location	Leaf Number	IIE, mm	DLPE, mm
X = -10 cm	16B	0.3	0.31
	20B	1.6	1.55
	23B	-0.5	-0.51
	31B	0.1	0.11
	33B	0.4	0.42
	34B	0.5	0.49
	37B	-0.4	-0.42
	44B	-0.2	-0.18
	45B	2.3	2.37
	50B	-0.5	-0.50
X = 0 cm	13A	0.8	0.76
	14A	-0.2	-0.19
	19A	-1.2	-1.20
	22A	-0.1	-0.09
	27A	2.7	2.66
	31A	0.1	0.11
	39A	-0.8	-0.79
	44A	0.2	0.22
	45A	-0.3	-0.28
	46A	-1.8	-1.81

DLPE = detected leaf positional error; IIE = intentionally introduced error.

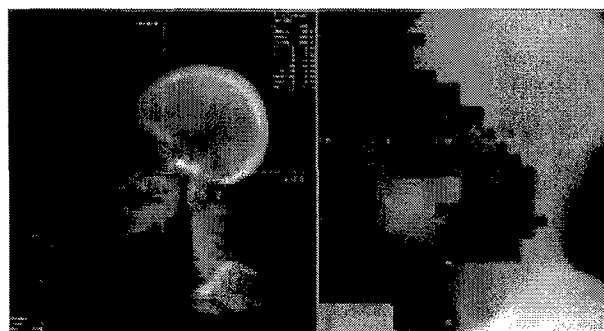


FIGURE 2-25. Top: A left digitally reconstructed radiograph *left* with the field boundary of the IMRT field. A double-exposure portal image for the same field is shown on the right. Bottom: An anterior posterior digitally reconstructed radiograph *left* with the field boundary of the IMRT field. A double-exposure portal image for the same field is shown on the right. These images are used by physicians to verify the maximum extent of the IMRT treatment fields.

calculation of the dose or MU at a point based on primitive machine data is recommended by AAPM Task Group 40.¹⁴¹ Because the fluence of a uniform or wedged field is spatially correlated, information of the dose at a point can, in principle, be used to estimate the dose in other points provided that the off-axis information is known. However, this is not the case for intensity-modulated fields because the weights of the beamlets across a field are independent. The correctness of the dose at a spatial point warrants, at most, only the correctness of the beamlets passing through or nearby that point. To validate an IMRT treatment plan, the spatial distributions of the beamlets must be verified in addition to the point dosimetric check.

In practice, the above two tasks can be achieved by the verification of point dose(s) and fluence maps. The fluence map of an incident beam is usually normalized to the maximum beamlet weight in the beam. For a given intensity-modulated field, the verification of the fluence map or beamlet correlation ensures the correctness of the doses at other points once the dose(s) at one or more points inside the field is examined. Together with the point doses, they provide information on the integrity of the IMRT fields.

We now discuss how to efficiently carry out the two types of tests. Obviously, the most robust method is to measure the point doses and fluence maps to validate an IMRT plan. As depicted as the dashed lines in Figure 2-26, the approach checks both planning and delivery. Its drawback is that an intensive effort is needed to carry out the measurement for each field or patient. Alternatively, one can separate the QA of the delivery and planning systems, as illustrated by the solid lines at the bottom. Although QA of the delivery system is imperative, its goal should be practically achievable by periodical checks rather than actual measurement before each patient's treatment. The division of IMRT QA into machine QA and patient-specific QA allows us to check the integrity of an IMRT treatment plan by using computer

calculation, simplifying the pretreatment QA. In fact, the same philosophy has been used in 3DCRT over the years, in which a manual calculation is often used instead of actual point dose measurement to validate the patient-specific MU settings.

Algorithms to perform the independent point dose and fluence map calculations for IMRT have been reported recently.¹⁵⁷⁻¹⁶² Here a general formalism for the IMRT point dose check used at Stanford University Hospital (the software, *IMSure*, has been commercialized by Prodigm Inc., Chico, CA) is described. In this approach, the dose at an arbitrary spatial point is expressed as a summation of the contributions from all of the beamlets with the amplitude of each beamlet modulated by a dynamic modulation factor. The dynamic modulation factor represents the fractional time that the beamlet is "open" during the dynamic delivery process and can be computed once the MLC leaf sequences are known.¹⁵⁹ The dose at a point (x, y, z) is written as

$$D(x, y, z) = MU \sum_m C_m D_m^0 \quad (18)$$

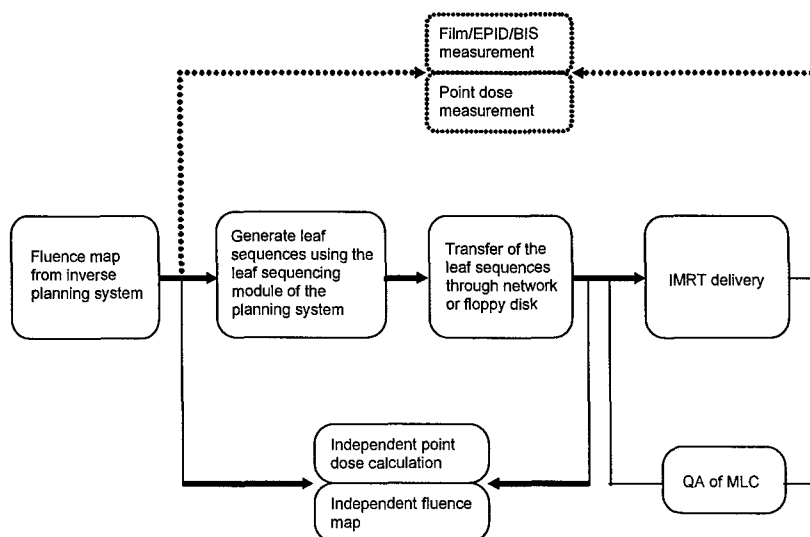
where the D_m^0 is the dose contribution to the calculation point per MU from the m -th beamlet when it is open, MU is the total monitor unit, and C_m is the dynamic modulation factor. When the MLC leaf transmission and head scatter effects are taken into account, C_m can be calculated by¹⁶⁰

$$C_m = \sum_k [Sc_{m,k} + \alpha Sc'_k (1 - \delta_{m,k})] f_k \quad (19)$$

with

$$\delta_{m,k} = \begin{cases} 1 & \text{if } m \in A_k \\ 0 & \text{if } m \notin A_k \end{cases} \quad (20)$$

FIGURE 2-26. Intensity-modulated radiation therapy (IMRT) plan validation process. The fluence map/point dose verification is depicted by the dashed line on the top. The computer-based approach is outlined as the solid lines at the bottom of the figure. BIS = beam imaging system; EPID = electronic portal imaging device; MLC = multileaf collimator; QA = quality assurance.



where f_k is the fractional MU of the k -th segment and A_k is the radiation field shape of the k -th segment. $Sc_{m,k}$ is the head scatter factor of the beamlet m in the k -th segment, Sc' is the head scatter factor for the rectangular field defined by the jaws, and α is the average transmission factor. The head scatter factor $Sc_{m,k}$ for each beamlet in a segment is calculated using the three-source model described earlier.¹²⁴

Computer verification of the fluence maps or the MLC leaf sequences can be done similarly. The software reads in the leaf sequences and simulates the motion of the MLC leaves.¹⁶³ The computed fluence map is then compared quantitatively with the intended map from the treatment planning system. A set of predefined QA indices are introduced to measure the "closeness" between the computed and the reference maps. The implication of the simulation is twofold. By comparing the recalculated fluence map with that from the planning system, it examines the functionality of the leaf sequencer of the planning system and ensures that the leaf sequence is executable and correct. It can also detect possible errors that occur during the transfer process of the leaf sequence file from the planning computer to the MLC workstation. The goal of the simulation is to warrant that, assuming that a rigorous independent QA of the MLC system has been performed so that the dynamic MLC can accurately execute the instruction of a leaf sequence file, the execution of the leaf sequence will generate the desired fluence map should it pass the simulation test.

Because of the simplicity and reliability of computer-based IMRT plan validation, it becomes clinically practical to enforce QA of the point doses and fluence maps on an individual patient or field basis. Furthermore, the method is valid for both step-and-shoot and dynamic deliveries. The utility of the computer verification has been demonstrated by the many clinical IMRT cases at many institutions, and its widespread use should simplify the QA procedure. However, it is important to keep in mind that experimental measurement is the only reliable source for IMRT plan validation. Any computer-based validation tool must be validated by experimental means before its clinical use.

Special IMRT Techniques and Machine Limitations

Concurrent Boost

One of the advantages of IMRT is its ability to deliver different dose levels to different regions simultaneously so that target volumes with different prescription doses can be planned and delivered together (see Chapter 18.7, "Simultaneous Integrated Boost: Emerging Technology").^{96,164} This approach has several potential advantages. Besides the efficiency of planning and delivery with a single plan, the resulting dose distribution can be more optimal. The conventional sequential boost strategy employs two or more independent plans in

which the initial fields cover the elective regions and smaller boost fields focus on the primary target. The boost dose is often limited by the tolerances of nearby OAR, which have been given a significant amount of radiation. If planned simultaneously using IMRT, it is possible to distribute the dose evenly among fractions, and the system also has a greater degree of freedom to optimize the intensity among many beams. There are biologic advantages as well; for example, the shortened treatment course and increased dose per fraction to primary tumors can often be translated into a higher biologically equivalent dose, thus increasing the probability of local control.

Treatment of Large Tumors

The treatment of large tumors necessitates the use of large treatment fields. Depending on the implementation of the MLC by the linear accelerator vendors, the maximum field size formed by dynamic MLC may be different from those imposed by the collimators (jaws). Typically, the maximum field size is much smaller. For example, in the Varian MLC, the jaws and the MLC carriages do not move with the leaves. The leaf length in the current model of the MLC is 14.5 cm (projected at the isocenter). Given that each leaf pair must travel from the left boundary to the right boundary of the beam aperture and the back end of any leaf cannot travel past the edge of the jaw, the maximum width of the field aperture that can be accommodated in one sweep of leaves is also limited to 14.5 cm. The maximum IMRT field size that can be delivered in one sweep is 40×14.5 cm for an 80-leaf MLC (or a 120-leaf MLC) or 26×14.5 cm for a 52-leaf MLC.

To treat a target volume wider than 14.5 cm, an incident field must be split into two or more subfields unless some special techniques are used.¹⁶⁵ A simple step "break" in the middle, as is usually done for static treatments with MLC, may be implemented. Although this is certainly feasible, it could lead to field matching problems because uncertainties in patient setup and leaf positioning may cause undesirable hot or cold spots in the junctioning region. Given that the intensity varies across the field in IMRT anyway, it is natural to consider splitting the beam into components with overlap between them having variable intensity in the overlap region. A simple dynamic "feathering" technique for splitting large fields has been proposed by Wu and colleagues.¹⁶⁶ In this method, the intensity-modulated field is divided into two (or more) components. The components overlap each other, and the intensity gradually decreases in the overlap region for one component and increases for the other. The sum of intensities remains the same as for the original field. Each component is delivered using the sweeping window technique with the dynamic MLC. This method provides a smooth transition from one field component to the next, thereby eliminating the field junction problems. The dynamic feathering technique may also be applied to split large static fields to minimize the junction problem.^{94,167} The feathering technique has been

extended to treat the whole abdomen area, in which splitting into more than two beams in the leaf motion direction may be necessary. Also, field sizes larger than 40 cm may be required in the cephalad-caudad direction, leading to the use of multiple isocenters, and feathering (not splitting) is helpful.¹⁶⁸

Dose Matching of an IMRT Plan with a 3DCRT or an IMRT Plan

One of the important problems in RT of breast cancer, Hodgkin's disease, head and neck cancer, and cervical cancer is the matching of an IMRT dose distribution for the treatment of part of the target volume(s) with a conventional 3DCRT or IMRT plan for the treatment of a different portion of target volume(s). Ideally, dose optimization of the second part should take into account the existing dose from the previous plan to optimally match the two dose distributions.^{94,169}

The two treatment plans that need to be matched are generally produced sequentially. The first plan used for treating part of the tumor volume(s) is obtained with the consideration of the second plan. To reduce the sensitivity of potential setup errors, an attempt needs to be made to "blur" the penumbra or dose gradient in the direction perpendicular to the matchline. Specifically, instead of a sharp dose gradient, the dose is allowed to extend by an additional 1.5 to 5 cm in the direction perpendicular to the matchline. In this transition region, the dose is forced to fall off linearly. The overlap is generally determined by the desired sensitivity against setup error. After the first plan is done with the extended transitional dose gradient region, the second IMRT plan is optimized with consideration of the existing doses of the first plan. The goal of the second dose optimization is to obtain an IMRT plan that yields a uniform composite dose distribution in the target volume(s) and (including the transitional regions) while sparing the sensitive structures. The approach takes advantage of the state-of-the-art intensity modulation and dose optimization techniques and provides an effective solution to the timely clinical problem of IMRT dose matching. In addition to better dose uniformity in the target volumes in the matchline region, it reduces the sensitivity of the doses to setup uncertainties in the matchline region. The technique is not yet available in commercial planning systems but should be implemented in the near future.

Radiation Protection Issues

Generally speaking, IMRT tends to use more beams (than traditional approaches) to conform the isodose curves to the shape of the tumor volume. As a consequence, a larger volume of normal tissue is exposed to lower doses as opposed to a smaller volume of normal tissue irradiated by higher doses in 3DCRT. In addition, the number of MUs is often increased by a factor of 2 to 3 owing to dynamic intensity

modulation, increasing the total-body exposure, which may increase the risk of secondary malignancies.^{170,171} Hall and Wu theoretically compared IMRT and 3DCRT and suggested that both factors tend to increase the risk of secondary cancers.¹⁷² Altogether, IMRT is likely to almost double the incidence of secondary malignancies compared with conventional RT (from about 1 to 1.75% for patients surviving 10 years or more). The risk may be larger for patients with longer survival rates (and for younger patients), but the ratio should remain the same (see Chapter 3 "Radiobiology of IMRT" and Chapter 30 "Pros and Cons of IMRT").

Reduction of the number of segments using more advanced dose optimization techniques and/or appropriate shielding of the treatment room are crucial to reduce the potential risks to hospital personnel. The National Council on Radiation Protection and Measurements has developed an empiric method for designing shielding against ionizing radiation that will protect workers and the general public from harmful radiation exposures.^{173,174} These methods have been used for several decades, and additional information that can be used in conjunction with these methods has since been published.¹⁷⁵⁻¹⁷⁷ A thorough study of IMRT shielding design has been presented by Mutic and colleagues¹⁷⁸ and Low.¹⁷⁹

Summary

Institutions worldwide are attempting or planning to integrate IMRT technology into their clinics. Before IMRT implementation, it is important to understand the physical principles behind the overall process of inverse planning and dynamic deliveries. This will help in making better decisions regarding which system best suits each clinical environment and facilitates the implementation process. The efficiency and quality of IMRT treatment depend on many factors. At this point, it seems that timely developments of inverse planning and QA techniques are highly desirable to make IMRT a truly superior and robust treatment modality. With these advancements, it is anticipated that IMRT will provide improved dose distributions with less effort in treatment planning, delivery, and verification.

Acknowledgments

We would like to thank J. G. Li, A. Pugachev, S. Crooks, C. Cotrutz, J. Lian, S. Hunjan, Z. Shou, E. Schreibmann, Y. Chen, G. Luxton, T. Pawlicki, Q. Le, S. Hancock, C. King, I. Gibbs, and B. Loo for many useful discussions. We also wish to acknowledge the support from the American Cancer Society (RSG-01-022-01-CCE), National Cancer Institute (5R01CA098523-02), Department of Defense (DAMD17-03-1-0023 and DAMD17-03-1-0019), and Vadasz Family Foundation. Last but not least, we wish to thank International Journal of Radiation Oncology, Biology, Physics, Medical Physics, and Physics in Medicine and Biology for permis-

sion to use their copyrighted materials.

References

1. Webb S. Intensity-modulated radiation therapy. Bristol (UK): Institute of Physics Publishing; 2001.
2. Webb S. Historical perspective on IMRT. In: Palta JR, Mackie TR, editors. AAPM intensity-modulated radiation therapy: the state of the art. Madison (WI): Medical Physics Publishing; 2003. p. 1–24.
3. Verhey LJ. Comparison of three-dimensional conformal radiation therapy and intensity-modulated radiation therapy systems. *Semin Radiat Oncol* 1999;9:78–98.
4. Yu CX. Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy. *Phys Med Biol* 1995;40:1435–49.
5. Yu CX, Li XA, Ma L, et al. Clinical implementation of intensity-modulated arc therapy. *Int J Radiat Oncol Biol Phys* 2002;53:453–63.
6. Wong E, Chen JZ, Greenland J. Intensity-modulated arc therapy simplified. *Int J Radiat Oncol Biol Phys* 2002;53:222–35.
7. Duthoy W, De Gersem W, Vergote K, et al. Whole abdominopelvic radiotherapy (WAPRT) using intensity-modulated arc therapy (IMAT): first clinical experience. *Int J Radiat Oncol Biol Phys* 2003;57:1019–32.
8. Crooks SM, Wu X, Takita C, et al. Aperture modulated arc therapy. *Phys Med Biol* 2003;48:1333–44.
9. Crooks S, King C, Pawlicki T, et al. Towards an optimal conformal technique for prostate radiotherapy: a dosimetric comparison between 3D CRT, IMRT, and IMAT. 2002. [In press].
10. Yang Y, Xing L. Clinical knowledge-based IMRT plan optimization. [Submitted]
11. Niemierko A. Reporting and analyzing dose distributions: a concept of equivalent uniform dose. *Med Phys* 1997;24:103–10.
12. Wu Q, Mohan R, Niemierko A, et al. Optimization of intensity-modulated radiation therapy plans based on the equivalent uniform dose. *Int J Radiat Oncol Biol Phys* 2002;52:224–35.
13. Thieke C, Bortfeld T, Niemierko A, et al. From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning. *Med Phys* 2003;30:2332–42.
14. Brahme A. Optimized radiation therapy based on radiobiological objectives. *Semin Radiat Oncol* 1999;9:35–47.
15. Niemierko A, Goitein M. Modeling of normal tissue response to radiation: the critical volume model. *Int J Radiat Oncol Biol Phys* 1993;25:135–45.
16. Niemierko A. Radiobiological models of tissue response to radiation in treatment planning systems. *Tumori* 1998;84:140–3.
17. Schultheiss TE, Orton CG, Peck RA. Models in radiotherapy: volume effects. *Med Phys* 1983;10:410–5.
18. Schultheiss TE, Orton CG. Models in radiotherapy: definition of decision criteria. *Med Phys* 1985;12:183–7.
19. Martel MK, Ten Haken RK, Hazuka MB, et al. Estimation of tumor control probability model parameters from 3-D dose distributions of non-small cell lung cancer patients. *Lung Cancer* 1999;24:31–7.
20. Martel MK. NTCP modeling for normal lung and tumors: analysis of clinical data. 3D conformal and intensity modulated radiation therapy: physics & clinical applications. Middleton (WI): Advanced Medical Physics Publishing; 2001. p. 489.
21. Kwa SL, Lebesque JV, Theuvs JC, et al. Radiation pneumonitis as a function of mean lung dose: an analysis of pooled data of 540 patients. *Int J Radiat Oncol Biol Phys* 1998;42:1–9.
22. Mohan R, Wang X, Jackson A, et al. The potential and limitations of the inverse radiotherapy technique. *Radiother Oncol* 1994;32:232–48.
23. Wang XH, Mohan R, Jackson A, et al. Optimization of intensity-modulated 3D conformal treatment plans based on biological indices. *Radiother Oncol* 1995;37:140–52.
24. Kallman P, Lind BK, Brahme A. An algorithm for maximizing the probability of complication-free tumour control in radiation therapy. *Phys Med Biol* 1992;37:871–90.
25. Xing L, Li JG, Donaldson S, et al. Optimization of importance factors in inverse planning. *Phys Med Biol* 1999;44:2525–36.
26. Xing L, Chen GTY. Iterative algorithms for inverse treatment planning. *Phys Med Biol* 1996;41:2107–23.
27. Xing L, Hamilton RJ, Spelbring D, et al. Fast iterative algorithms for three-dimensional inverse treatment planning. *Med Phys* 1998;25:1845–9.
28. Brooks RA, DeChiro G. Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Phys Med Biol* 1976;21:689–732.
29. Jones L, Hoban P. A method for physically based radiotherapy optimization with intelligent tissue weight determination. *Med Phys* 2002;29:26–37.
30. Luenberger D. Introduction to linear and nonlinear programming. Reading (MA): Addison-Wesley; 1973.
31. Kirkpatrick S, Gelatt C, Vecchi M. Optimization by simulated annealing. *Science* 1983;220:671–80.
32. Metropolis N, Rosenbluth A, Rosenbluth M, et al. Equation of state calculation by fast computing machines. *J Chem Phys* 1953;21:1087–91.
33. Rosen, II, Lam KS, Lane RG, et al. Comparison of simulated annealing algorithms for conformal therapy treatment planning. *Int J Radiat Oncol Biol Phys* 1995;33:1091–9.
34. Webb S. Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator. *Phys Med Biol* 1991;36:1201–26.
35. Winkler G. Image analysis, random field and dynamic Monte Carlo methods. Berlin: Springer-Verlag; 1995.
36. Webb S. Optimization by simulated annealing of three-dimensional, conformal treatment planning for radiation fields defined by a multileaf collimator: II. Inclusion of two-dimensional modulation of the x-ray intensity. *Phys Med Biol* 1992;37:1689–704.
37. Bortfeld T, Burkhardt J, Boesecke R, et al. Methods of image reconstruction from projections applied to conformation radiotherapy. *Phys Med Biol* 1990;35:1423–34.
38. Holmes T, Mackie TR. A filtered backprojection dose calculation method for inverse treatment planning. *Med Phys* 1994;21:303–13.

39. Pugachev AB, Boyer AL, Xing L. Beam orientation optimization in intensity-modulated radiation treatment planning. *Med Phys* 2000;27:1238–45.
40. Andrew HC, Hunt BR. Image restoration. Upper Saddle River (NJ): Prentice Hall; 1977.
41. Starkschall G. A constrained least-squares optimization method for external beam radiation therapy treatment planning. *Med Phys* 1984;11:659–65.
42. Starkschall G, Eifel PJ. An interactive beam-weight optimization tool for three-dimensional radiotherapy treatment planning. *Med Phys* 1992;19:155–63.
43. Lee EK, Fox T, Crocker I. Optimization of radiosurgery treatment planning via mixed integer programming. *Med Phys* 2000;27:995–1004.
44. Bednarz G, Michalski D, Houser C, et al. The use of mixed-integer programming for inverse treatment planning with pre-defined field segments. *Phys Med Biol* 2002;47:2235–45.
45. Langer M, Morrill SS, Brown R, et al. A comparison of mixed integer programming and fast simulated annealing for optimizing beam weights in radiation therapy. *Med Phys* 1996;23:957–964.
46. Lee EK, Zaider M. Intraoperative dynamic dose optimization in permanent prostate implants. *Int J Radiat Oncol Biol Phys* 2003;56:854–61.
47. Hou Q, Wang J, Chen Y, et al. An optimization algorithm for intensity modulated radiotherapy—the simulated dynamics with dose-volume constraints. *Med Phys* 2003;30:61–8.
48. Langer M, Brown R, Morrill S, et al. A generic genetic algorithm for generating beam weights. *Med Phys* 1996;23:965–71.
49. Ezzell GA, Gaspar L. Application of a genetic algorithm to optimizing radiation therapy treatment plans for pancreatic carcinoma. *Med Dosim* 2000;25:93–7.
50. Wu X, Zhu Y. A mixed-encoding genetic algorithm with beam constraint for conformal radiotherapy treatment planning. *Med Phys* 2000;27:2508–16.
51. Cotrutz C, Xing L. Segment-based dose optimization using a genetic algorithm. *Phys Med Biol* 2003;48:2987–98.
52. Xing L, Li JG, Pugachev A, et al. Estimation theory and model parameter selection for therapeutic treatment plan optimization. *Med Phys* 1999;26:2348–58.
53. Lian J, Cotrutz C, Xing L. Therapeutic treatment plan optimization with probabilistic dose prescription. *Med Phys* 2003;30:655–66.
54. Winkler RL. An introduction to bayesian inference and decision. New York: Holt, Rinehart & Winston; 1972.
55. Llacer J. Inverse radiation treatment planning using the dynamically penalized likelihood method. *Med Phys* 1997;24:1751–64.
56. Llacer J, Solberg TD, Promberger C. Comparative behaviour of the dynamically penalized likelihood algorithm in inverse radiation therapy planning. *Phys Med Biol* 2001;46:2637–63.
57. Wu X, Zhu Y. A maximum-entropy method for the planning of conformal radiotherapy. *Med Phys* 2001;28:2241–6.
58. Wu X, Zhu Y. A neural network regression model for relative dose computation. *Phys Med Biol* 2000;45:13–22.
59. Li RP, Yin FF. Optimization of inverse treatment planning using a fuzzy weight function. *Med Phys* 2000;27:691–700.
60. Ezzell G. Clinical implementation of IMRT treatment planning. In: Palta JR, Mackie TR, editors. Intensity-modulated radiation therapy: the state of the art. Colorado Springs (CO): Medical Physics Publishing; 2003. p. 475–94.
61. Xing L, Lin Z, Donaldson SS, et al. Dosimetric effects of patient displacement and collimator and gantry angle misalignment on intensity modulated radiation therapy. *Radiother Oncol* 2000;56:97–108.
62. Manning MA, Wu Q, Cardinale RM, et al. The effect of setup uncertainty on normal tissue sparing with IMRT for head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 2001;51:1400–9.
63. International Commission on Radiation Units and Measurements. Recording and reporting photon beam therapy. Report 50. Washington (DC): International Commission on Radiation Units and Measurements; 1993.
64. Bortfeld T, Schlegel W. Optimization of beam orientations in radiation therapy: some theoretical considerations. *Phys Med Biol* 1993;38:291–304.
65. Pugachev A, Li JG, Boyer AL, et al. Role of beam orientation optimization in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 2001;50:551–60.
66. Pugachev A, Xing L, Boyer AL. Beam orientation optimization in IMRT: to optimize or not to optimize? Presented at the XII International Conference on the Use of Computers in Radiation Therapy; 2000; Heidelberg, Germany.
67. Soderstrom S, Brahme A. Selection of suitable beam orientations in radiation therapy using entropy and Fourier transform measures. *Phys Med Biol* 1992;37:911–24.
68. Rowbottom CG, Nutting CM, Webb S. Beam-orientation optimization of intensity-modulated radiotherapy: clinical application to parotid gland tumours. *Radiother Oncol* 2001;59:169–77.
69. Djajaputra D, Wu Q, Wu Y, et al. Algorithm and performance of a clinical IMRT beam-angle optimization system. *Phys Med Biol* 2003;48:3191–212.
70. Stein J, Mohan R, Wang XH, et al. Number and orientations of beams in intensity-modulated radiation treatments. *Med Phys* 1997;24:149–60.
71. Pugachev A, Xing L. Computer assisted beam orientation selection in IMRT. *Phys Med Biol* 2001;46:2467–76.
72. Pugachev A, Xing L. Incorporating prior knowledge into beam orientation optimization. *Int J Radiat Oncol Biol Phys* 2002;54:1565–74.
73. Xing L, Pugachev A, Li JG, et al. A medical knowledge based system for the selection of beam orientations in IMRT. *Int J Radiat Oncol Biol Phys* 1999;45:246.
74. Schreiber E, Xing L. Feasibility study of beam orientation class-solutions for prostate IMRT. *Medical Physics*; 2004. [In press]
75. Kay S. Fundamentals of statistical signal processing: estimation theory. Upper Saddle River (NJ): Prentice Hall; 1993.
76. Wu Q, Mohan R. Algorithms and functionality of an intensity modulated radiotherapy optimization system. *Med Phys* 2000;27:701–11.
77. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys* 1991;21:109–22.
78. Eisbruch A, Ten Haken RK, Kim HM. Dose, volume, and function relationships in parotid salivary glands following conformal and intensity-modulated irradiation of head and neck cancer. *Int J Radiat Oncol Biol Phys* 1999;45:577–87.

79. Lian J, Xing L. Incorporating model parameter uncertainty into inverse treatment planning. *Med Phys* 2004. [In press]
80. Chen Y, Michalski D, Houser C. A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy. *Phys Med Biol* 2002;47:1647–58.
81. Yu Y, Zhang JB, Cheng G, et al. Multi-objective optimization in radiotherapy: applications to stereotactic radiosurgery and prostate brachytherapy. *Art Intel Med* 2000;19:39–51.
82. Cotrutz C, Lahanas M, Kappas C, et al. A multiobjective gradient-based dose optimization algorithm for external beam conformal radiotherapy. *Phys Med Biol* 2001;46:2161–75.
83. Lahanas M, Schreiber E. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. *Phys Med Biol* 2003;48:2843–71.
84. Bortfeld T. Physical optimization. In: Palta JR, Mackie TR, editors. *AAPM intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 51–76.
85. De Gersem W, Claus F, De Wagter C, et al. Leaf position optimization for step-and-shoot IMRT. *Int J Radiat Oncol Biol Phys* 2001;51:1371–88.
86. Shepard DM, Earl MA, Li XA, et al. Direct aperture optimization: a turnkey solution for step-and-shoot IMRT. *Med Phys* 2002;29:1007–18.
87. Cotrutz C, Xing L. Using voxel-dependent importance factors for interactive DVH-based dose optimization. *Phys Med Biol* 2002;47:1659–69.
88. Cotrutz C, Xing L. IMRT dose shaping using regionally variable penalty scheme. *Med Phys* 2003;30:544–51.
89. Rowbottom CG, Webb S, Oldham M. Improvements in prostate radiotherapy from the customization of beam directions. *Med Phys* 1998;25:1171–9.
90. Das S, Cullip T, Tracton G, et al. Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization. *Int J Radiat Oncol Biol Phys* 2003;55:215–24.
91. Bedford JL, Webb S. Elimination of importance factors for clinically accurate selection of beam orientations, beam weights and wedge angles in conformal radiation therapy. *Med Phys* 2003; 30:1788–881.
92. Pugachev A, Xing L. Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. *Int J Radiat Oncol Biol Phys* 2001;51:1361–70.
93. Holt R, Xing L. Salvage of suboptimal prostate seed implants using IMRT. *Med Phys* 2001;28:1308.
94. Li JG, Xing L, Boyer AL, et al. Matching photon and electron fields with dynamic intensity modulation. *Med Phys* 1999;26:2379–84.
95. Chao KS, Majhail N, Huang CJ, et al. Intensity-modulated radiation therapy reduces late salivary toxicity without compromising tumor control in patients with oropharyngeal carcinoma: a comparison with conventional techniques. *Semin Radiat Oncol* 2002;12(1 Suppl 1):20–5.
96. Wu Q, Mohan R, Morris M, et al. Simultaneous integrated boost intensity-modulated radiotherapy for locally advanced head-and-neck squamous cell carcinomas. I: dosimetric results. *Int J Radiat Oncol Biol Phys* 2003;56:573–85.
97. Mundt AJ, Lujan AE, Rotmensch J, et al. Intensity-modulated whole pelvic radiotherapy in women with gynecologic malignancies. *Int J Radiat Oncol Biol Phys* 2002;52:1330–7.
98. Forster KM, Smythe WR, Starkschall G, et al. Intensity-modulated radiotherapy following extrapleural pneumonectomy for the treatment of malignant mesothelioma: clinical implementation. *Int J Radiat Oncol Biol Phys* 2003;55:606–16.
99. Hanley J, Debois MM, Mah D, et al. Deep inspiration breath-hold technique for lung tumors: the potential value of target immobilization and reduced lung density in dose escalation. *Int J Radiat Oncol Biol Phys* 1999;45:603–11.
100. Murphy MJ, Martin D, Whyte R, et al. The effectiveness of breath-holding to stabilize lung and pancreas tumors during radiosurgery. *Int J Radiat Oncol Biol Phys* 2002;53:475–82.
101. Bortfeld T, Jokivarsi K, Goitein M, et al. Effects of intra-fraction motion on IMRT dose delivery: statistical analysis and simulation. *Phys Med Biol* 2002;47:2203–20.
102. Ramsey CR, Scaperroth D, Arwood D, et al. Clinical efficacy of respiratory gated conformal radiation therapy. *Med Dosim* 1999;24:115–9.
103. Wong JW, Sharpe MB, Jaffray DA, et al. The use of active breathing control (ABC) to reduce margin for breathing motion. *Int J Radiat Oncol Biol Phys* 1999;44:911–9.
104. Vedam SS, Keall PJ, Kini VR, et al. Determining parameters for respiration-gated radiotherapy. *Med Phys* 2001;28:2139–46.
105. Keall PJ, Kini VR, Vedam SS, et al. Motion adaptive x-ray therapy: a feasibility study. *Phys Med Biol* 2001;46:1–10.
106. Kini VR, Vedam SS, Keall PJ, et al. Patient training in respiratory-gated radiotherapy. *Med Dosim* 2003;28:7–11.
107. Ling CC, Humm J, Larson S, et al. Towards multidimensional radiotherapy (MD-CRT): biological imaging and biological conformality. *Int J Radiat Oncol Biol Phys* 2000;47:551–60.
108. Rosenman J. Incorporating functional imaging information into radiation treatment. *Semin Radiat Oncol* 2001;11:83–92.
109. Xing L, Cotrutz C, Hunjan S, et al. Inverse planning for functional image-guided IMRT. *Phys Med Biol* 2002;47:3567–78.
110. Alber M, Nusslin F. An objective function for radiation treatment optimization based on local biological measures. *Phys Med Biol* 1999;44:479–93.
111. Bortfeld T, Boyer AL, Schlegel W, et al. Realization and verification of three-dimensional conformal radiotherapy with modulated fields. *Int J Radiat Oncol Biol Phys* 1994;30:899–908.
112. Xia P, Verhey LJ. Multileaf collimator leaf sequencing algorithm for intensity modulated beams with multiple static segments. *Med Phys* 1998;25:1424–34.
113. Crooks SM, McAviney LF, Robinson DF, et al. Minimizing delivery time and monitor units in static IMRT by leaf-sequencing. *Phys Med Biol* 2002;47:3105–16.
114. Siochi RA. Minimizing static intensity modulation delivery time using an intensity solid paradigm. *Int J Radiat Oncol Biol Phys* 1999;43:671–80.
115. Langer M, Thai V, Papiez L, et al. Improved leaf sequencing reduces segments or monitor units needed to deliver IMRT using multileaf collimators: the influence of irradiation and packaging on the quality of prepacked vegetables. *Med Phys* 2001;28:2450–8.

116. Beavis AW, Ganney PS, Whitton VJ, et al. Optimization of the step-and-shoot leaf sequence for delivery of intensity modulated radiation therapy using a variable division scheme. *Phys Med Biol* 2001;46:2457-65.
117. Convery DJ, Rosenbloom ME. Generation of intensity modulated fields by dynamic collimation. *Phys Med Biol* 1992;37:1359-74.
118. Svensson R, Kallman P, Brahme A. An analytical solution for the dynamic control of multileaf collimators. *Phys Med Biol* 1994;39:37-61.
119. Spirou SV, Chui CS. Generation of arbitrary intensity profiles by dynamic jaws or multileaf collimators. *Med Phys* 1994;21:1031-41.
120. Dai JR, Hu YM. Intensity-modulation radiotherapy using independent collimators: an algorithm study. *Med Phys* 1999;26:2562-70.
121. Beavis AW, Ganney PS, Whitton VJ, et al. Slide and shoot: a new method for MLC delivery of IMRT. In: *Proceedings of the XIII International Conference on the Use of Computers in Radiation Therapy*, Heidelberg, Germany, May 2000.
122. Beavis A, Ganney P, Whitton V, Xing L. Slide and shoot: a new method for MLC delivery of IMRT. *The Use of Computers in Radiation Therapy*, Heidelberg, Germany, 2000.
123. Yang Y, Xing L. Incorporating leaf transmission and header scatter corrections into MLC leaf sequences for IMRT. *Int J Radiat Oncol Biol Phys* 2003;55:1121-34.
124. Yang Y, Xing L, Boyer A, et al. A three-source model for the calculation of head scatter factors. *Med Phys* 2002;29:2024-33.
125. Chen Y, Boyer AL, Ma CM. Calculation of x-ray transmission through a multileaf collimator. *Med Phys* 2000;27:1717-26.
126. Deng J, Pawlicki T, Chen Y, et al. The MLC tongue-and-groove effect on IMRT dose distributions. *Phys Med Biol* 2001;46:1039-60.
127. Hounsell AR. Monitor chamber backscatter for intensity modulated radiation therapy using multileaf collimators. *Phys Med Biol* 1998;43:445-54.
128. Chui CS, LoSasso T, Spirou S. Dose calculation for photon beams with intensity modulation generated by dynamic jaw or multileaf collimations. *Med Phys* 1994;21:1237-44.
129. Wang X, Spirou S, LoSasso T, et al. Dosimetric verification of intensity-modulated fields. *Med Phys* 1996;23:317-27.
130. van Santvoort JP, Heijmen BJ. Dynamic multileaf collimation without 'tongue-and-groove' underdosage effects. *Phys Med Biol* 1996;41:2091-105.
131. Webb S, Bortfeld T, Stein J, et al. The effect of stair-step leaf transmission on the 'tongue-and-groove problem' in dynamic radiotherapy with a multileaf collimator. *Phys Med Biol* 1997;42:595-602.
132. Dirx ML, Heijmen BJ, van Santvoort JP. Leaf trajectory calculation for dynamic multileaf collimation to realize optimized fluence profiles. *Phys Med Biol* 1998;43:1171-84.
133. Ma L, Boyer AL, Ma CM, et al. Synchronizing dynamic multileaf collimators for producing two-dimensional intensity-modulated fields with minimum beam delivery time. *Int J Radiat Oncol Biol Phys* 1999;44:1147-54.
134. Galvin JM, Smith AR, Lally B. Characterization of a multileaf collimator system. Comment in: *Int J Radiat Oncol Biol Phys* 1993 Jan 15;25(2):373-5. *Int J Radiat Oncol Biol Phys* 1993;25:181-92.
135. Sykes JR, Williams PC. An experimental investigation of the tongue and groove effect for the Philips multileaf collimator. *Phys Med Biol* 1998;43:3157-65.
136. IMRT Collaborative Working Group. Intensity-modulated radiotherapy: current status and issues of interest. *Int J Radiat Oncol Biol Phys* 2001;51:880-914.
137. Sharpe MB. Commissioning and quality assurance for IMRT treatment planning. In: Palta JR, Mackie TR, editors. *Intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 449-74.
138. Xia P, Chuang C. Patient-specific quality assurance in IMRT. In: Palta JR, Mackie TR, editors. *Intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 495-514.
139. Moran JM. Dosimetry metrology. In: Palta JR, Mackie TR, editors. *Intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 415-38.
140. Xing L, Curran B, Hill R, et al. Dosimetric verification of a commercial inverse treatment planning system. *Phys Med Biol* 1999;44:463-78.
141. Kutcher GJ, Coia L, Gillin M, et al. Comprehensive QA for radiation oncology: report of AAPM Radiation Therapy Committee Task Group 40. *Med Phys* 1994;21:581-618.
142. Fraass B, Doppke K, Hunt M, et al. American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. *Med Phys* 1998;25:1773-829.
143. Ezzell G, Galvin J, Low D, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT Subcommittee of the AAPM Radiation Therapy Committee. *Med Phys* 2003;30:2089-115.
144. Bortfeld TR, Kahler DL, Waldron TJ, et al. X-ray field compensation with multileaf collimators. *Int J Radiat Oncol Biol Phys* 1994;28:723-30.
145. Ma L, Boyer AL, Xing L, et al. An optimized leaf-setting algorithm for beam intensity modulation using dynamic multileaf collimators. *Phys Med Biol* 1998;43:1629-43.
146. Crooks S, Pugachev A, King C, et al. Examination of the effect of increasing the number of radiation beams on a radiation treatment plan. *Phys Med Biol* 2002;47:3485-501.
147. Van Dyk J, Purdy JA. Clinical implementation of technology and the quality assurance process. In: Van Dyk J, editor. *The modern technology of radiation oncology*. Madison (WI): Medical Physics Publishing; 1999. p. 19-51.
148. Chui CS, Spirou S, LoSasso T. Testing of dynamic multileaf collimation. *Med Phys* 1996;23:635-41.
149. LoSasso TJ. IMRT delivery system QA. In: Palta JR, Mackie TR, editors. *AAPM intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 561-91.
150. Herman MG, Balter JM, Jaffray DA, et al. Clinical use of electronic portal imaging: report of AAPM Radiation Therapy Committee Task Group 58. *Med Phys* 2001;28:712-37.
151. Chang J, Mueller K, Sidhu K, et al. Verification of multileaf collimator leaf positions using an electronic portal imaging device. *Med Phys* 2002;29:2913-24.

152. Samant SS, Zheng W, Parra NA, et al. Verification of multileaf collimator leaf positions using an electronic portal imaging device. *Med Phys* 2002;29:2900-12.
153. Vieira SC, Dirkx ML, Pasma KL, Heijmen BJ. Fast and accurate leaf verification for dynamic multileaf collimation using an electronic portal imaging device. *Med Phys* 2002;29:2034-40.
154. Yang Y, Xing L. Using the volumetric effect of a finite-sized detector for routine quality assurance of MLC leaf positioning. *Med Phys* 2003;30:433-41.
155. Yang Y, Xing L. Quantitative measurement of MLC leaf displacements using an electronic portal image device. *Phys Med Biol* 2004;49:1251-1533.
156. Chen Y, Xing L, Luxton G, et al. A multi-purpose quality assurance tool for MLC-based IMRT. ICCR, Heidelberg, Germany, May, 2000.
157. Boyer A, Xing L, Ma CM, et al. Theoretical considerations of monitor unit calculations for intensity modulated beam treatment planning. *Med Phys* 1999;26:187-95.
158. Kung J, Chen G. A monitor unit verification calculation in intensity modulated radiotherapy as a dosimetric quality assurance. *Med Phys* 2000;27:2226-30.
159. Xing L, Chen Y, Luxton G, et al. Monitor unit calculation for an intensity modulated photon field by a simple scatter-summation algorithm. *Phys Med Biol* 2000;45:N1-7.
160. Yang Y, Xing L, Li JL, et al. Independent dosimetric calculation with inclusion of head scatter and MLC transmission for IMRT. *Med Phys* 2003;30:2937-47.
161. Xing L, Yang Y, Li J, et al. Monitor unit calculation and plan validation for IMRT. In: Palta JR, Mackie TR, editors. *Intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 3567-78.
162. Watanabe Y. Point dose calculations using an analytical pencil beam kernel for IMRT plan checking. *Phys Med Biol* 2001;46:1031-8.
163. Xing L, Li JG. Computer verification of fluence maps in intensity modulated radiation therapy. *Med Phys* 2000;27:2084-92.
164. Scheffter TE, Kavanagh BD, Wu Q, et al. Technical considerations in the application of intensity-modulated radiotherapy as a concomitant integrated boost for locally-advanced cervix cancer. *Med Dosim* 2002;27:177-84.
165. Xing L, Yi BY, Li J, et al. Adaptive inverse planning with consideration of MLC field size constraint. Presented at the American Association of Physicists in Medicine Annual Meeting; 1999; Nashville, TN.
166. Wu Q, Arnfield M, Tong S, et al. Dynamic splitting of large intensity-modulated fields. *Phys Med Biol* 2000;45:1731-40.
167. Xing L, Boyer A, Kapp D, et al. Improving the matching of abutting photon fields by modulating photon beams. *Med Biol Eng Comp* 1997;35:921.
168. Hong L, Alektiar K, Chui C, et al. IMRT of large fields: whole-abdomen irradiation. *Int J Radiat Oncol Biol Phys* 2002;54:278-89.
169. Xing L, Yang Y, Li J, et al. Dose matching of an IMRT plan with an electron or 3D conformal treatment plan. Presented at the World Congress on Medical Physics and Biomedical Engineering; 2003 Aug 24-29; Sydney, Australia.
170. Followill D, Geis P, Boyer A. Estimates of whole-body dose equivalent produced by beam intensity modulated conformal therapy. *Int J Radiat Oncol Biol Phys* 1997;38:667-72.
171. Verellen D, Vanhavere F. Risk assessment of radiation-induced malignancies based on whole-body equivalent dose estimates for IMRT treatment in the head and neck region. *Radiother Oncol* 1999;53:199-203.
172. Hall EJ, Wu CS. Radiation-induced second cancers: the impact of 3D-CRT and IMRT. *Int J Radiat Oncol Biol Phys* 2003;56:83-8.
173. Radiation protection design guidelines for 0.1-100 MeV particle accelerator facilities. Vol. 51. National Council on Radiation Protection and Measurements; Washington (DC): 1977.
174. Structural shielding design and evaluation for medical use of x-rays and gamma rays of energies up to 10 MeV. Vol 49. National Council on Radiation Protection and Measurements. Washington (DC): 1976.
175. McGinley PH. Shielding techniques for radiation oncology facilities. 2nd ed. Madison (WI): Medical Physics Publishing; 2002.
176. McGinley HM, Miner MS. A history of radiation shielding of x-ray therapy rooms. *Health Phys* 1995;69:759-65.
177. Rodgers JE. Radiation therapy vault shielding calculation methods when IMRT and TBI procedures contribute. *J Appl Clin Med Phys* 2001;2:157-64.
178. Mutic S, Low DA, Klein EE, et al. Room shielding for intensity-modulated radiation therapy treatment facilities. *Int J Radiat Oncol Biol Phys* 2001;50:39-46.
179. Low D. Radiation shielding for IMRT. In: Palta JR, Mackie TR, editors. *AAPM intensity-modulated radiation therapy: the state of the art*. Madison (WI): Medical Physics Publishing; 2003. p. 401-14.